Generating an efficient remote collaboration environment using shared gaze

A thesis proposal presented
By
Sathya Kumar Barathan
To
The Human Interface Technology Laboratory NZ

In partial fulfilment of the requirements
for the degree of
Master's
In the Subject of
Human Interface Technology

University of Canterbury
Christchurch, Canterbury
July 2016

Senior supervisor – **Dr. Gun Lee**
Co-supervisor – **Prof. Mark Billinghurst, Prof. Robert Lindeman**

Generating an efficient remote collaboration environment using shared gaze

ABSTRACT

Visual sharing over long distance has become an everyday medium with the current advancements in technology. A person can simply wear a head-mounted display (HMD) such as the Google Glass and enable another person to view what they see on a remote desktop. My thesis tested an interface to see if video conferencing experience could be improved through different cues from a remote user. Hereby, making it simple for the person wearing the head-mount to communicate more efficiently in the video sharing process. The primary focus of this interface is to simplify the human-human interaction over video conferencing. As a part of creating this interface, I reviewed the literature of wearable devices, different forms of communication processes and AR-based advancements. As part of my thesis, I developed a prototype system to test the effects of different cues in a collaborative video-conferencing environment. In the experiment, a remote user instructed a local user by sharing their camera feed using a HMD to perform a set of actions using varied conditions. The experiment used a HMD with an attached camera for the local user and an eye tracker for the remote user. User studies, interviews and tests were performed in different stages to approve the adoption of the prototype. The results showed that the users preferred the mouse + voice and gaze + voice cues over voice only cues in the remote collaborative process over video conferencing. However, there was a trend showing that the participants were in favour of using gaze cues over the other two conditions considering the ease-of-use and clear communication.

Table of Contents

Listing of Figures

Listing of Tables

I DEDICATE MY THESIS TO MY FATHER WHO HAS BEEN MORE THAN A FRIEND, MENTOR AND HUGE SUPPORT THROUGHOUT MY LIFE.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my senior supervisor Dr. Gun Lee for being a mentor and supporting me throughout my thesis and providing his extensive knowledge while allowing me to work in my own way. I also wish to thank Prof. Rob Lindeman for sharing his expertise, sincere and valuable guidance and encouragement. I wish to thank my co-supervisor Prof. Mark Billinghurst for his insightful comments and for the assistance provided at all levels of my thesis. I would like to thank Ken Beckman for his valuable assistance in obtaining my scholarship and for various hurdles faced during the study. I also thank CRCSI for funding my scholarship for my course study. I also wish to thank my parents, friends, colleagues and staff at the HIT Lab NZ for all their support, exchange of knowledge and the wonderful experience of doing research there.

1

Introduction

This thesis expands on sharing a holistic visual experience in distant visual association, as in video conferencing. Long distance communication has several key elements: reciprocated voice recognition, visual learning and recognition of visual objects. Users should be able to share the same experience, regardless of their physical location. Broad research in contemporary developmental psychology has proven that this mutual immersion is one of the most important aspects in any communal interaction [1].

Previous research has extensively explored different forms of virtual collaboration, from remote media immersion to distributed immersive performance (DIP) [2]. Research in virtual collaboration is critical for engineering, applied sciences and working prototypes of immersive live environments. For example, one of the outcomes of using DIP is to create a remote collaborative environment where participants of a musical performance located in different regions can work together in a wider virtual space with high-fidelity audio and visual channels [2]. Real-time communication using shared gaze would be effective in enhancing remote collaboration [3].

1.1 Motivation

Early studies have proved that spoken communication is the most efficient way to collaborate remotely, as compared to email or other forms of communication. However, for collaborative tasks, voice-only interaction is more time-consuming than using visual and voice cues to communicate instructions [4]. Current studies are working on using visual interaction to bridge the gap in spatial referencing. This research promises to improve the communicative process during remote collaboration.

Several other joint communication systems that require mutual attention emphasise spatial referencing — to define a particular object's location and confirming it. Spatial referencing is one of the most important tasks during joint immersion. For instance, in a virtual crime scene [see Figure 1.1], the instructor is sitting in the control room and another person is seeing the crime scene using a head-mounted device. In Figure 1.1, the yellow circle represents the eye-gaze of the participant in the scene and the red dot (adjusted) reflects the instructor's gaze position. Neider et al. found that this collaborative experience is enhanced using shared eye gaze [5].

[Figure 1.1] An illustrated shared gaze simulation reflecting shared gaze in a two-way remote collaboration [5].

The value of gaze engagement is valuable in remote collaboration. Gaze association creates a shared focus on appropriate objects, as described by Bruner as joint attention [6]. Studies have shown that gaze transfer improves the conversational foundation in co-dependent settings. Brennan et al. [7] experimented with eyetracking and vocal communication while users explored sniper targets together in a street scene. They concluded that the partner's gaze was vital to complete the assigned mission [see Figure

1.2]. Expanding on their research on eye tracking in VR environment, this thesis explores on the efficiency of eye gaze in video conferencing systems on real world tasks with the help of head worn remote collaboration systems and gaze trackers.

[Figure 1.2] Two participants wearing an eye tracker from different rooms searching for targets using joint gaze and speech at different times [7]

1.2 Goal and Research Opportunities

Expanding from earlier research on gaze perception, affective computing, spatial referencing using gaze coordination and video conferencing, my thesis involved the following research opportunities:

- **Advancements in wearable technology** that emphasise AR displays and gaze tracking
- **A communication system** that uses the human eye to reduce the demands on speech
- **Eye tracking** for tasks cued in a remote user's virtual environment and **Spatial referencing** through gaze coordination in video conferencing

The primary goal of my research is to compare different cues (voice, mouse + voice, gaze + voice) from a remote user in a video conferencing based collaborative environment. The study used an interactive prototype in which a remote user guides their partner in performing a set of similar actions in a real-life scenario using these different cues. This is discussed in the detail in the following chapters.

1.3 Scope

There is a wide range of communication technologies, AR annotations and visual cues that can be used to virtually collaborate over video conferencing. In contrast to earlier research on sharing local worker's eye gaze [8], my research is more focussed on the remote user's eye gaze. Hence, the local user was simply following the instructions from the remote user achieve the assigned task. Nevertheless, the local user's feedback was measured in the qualitative analysis since the remote user's experience is affected by the accuracy of the cues and if they are able to receive the instructions well through their HMD during the experiment.

Based on previous research and interviews from industrial workers, I tried to identify the types of users who would benefit from a remote collaborative process. An elaborative description of potential users is included in section 3.1 of this thesis.

In order to ideate a high-fidelity prototype to compare the conditions in the collaborative process, I studied different tasks that could help estimate the value of different cues. For example, tasks such as building a LEGO set, assembling a puzzle and using a demo control panel were explored before finalising the user study.

In building the prototype, I examined the current market options for eye trackers and head-mounted displays before finalising my experiment set-up.

1.4 Thesis structure

Following this introduction to the research, I first present a summary of prior research on the applications of video conferencing in a collaborative environment, head-mounted displays, eye tracking and shared emotional experience (Chapter 2, Related work). I also summarise the elements of similar research that have guided this thesis.

Next, I discuss the design process and the pilot study (Chapter 3, Design). This includes results from brainstorming sessions and interviews, the study on users and test scenarios, and the limitations that were kept in mind before building the user experiment.

After analysing the results from the initial process and brainstorming sessions with my peers, I devised design solutions and concepts for low-fidelity prototypes.

In a chapter on prototype implementation, I elaborate on the hardware and software used in the construction of the prototype (Chapter 4).Evaluation purpose, experiment design, experiment procedure and the summarised results have been included in Chapter 5 (Evaluation). The findings are also compared with previous studies on similar topics in the Discussion section in the later chapters of this thesis (Chapter 6). Further discussion on results and future work has been elaborated in the final chapters (Chapter 7, Conclusion and Future Work).

This thesis also includes an Appendix consisting of the various visual designs, consent forms and questionnaires that were presented to the participants during the research process.

2

Related work

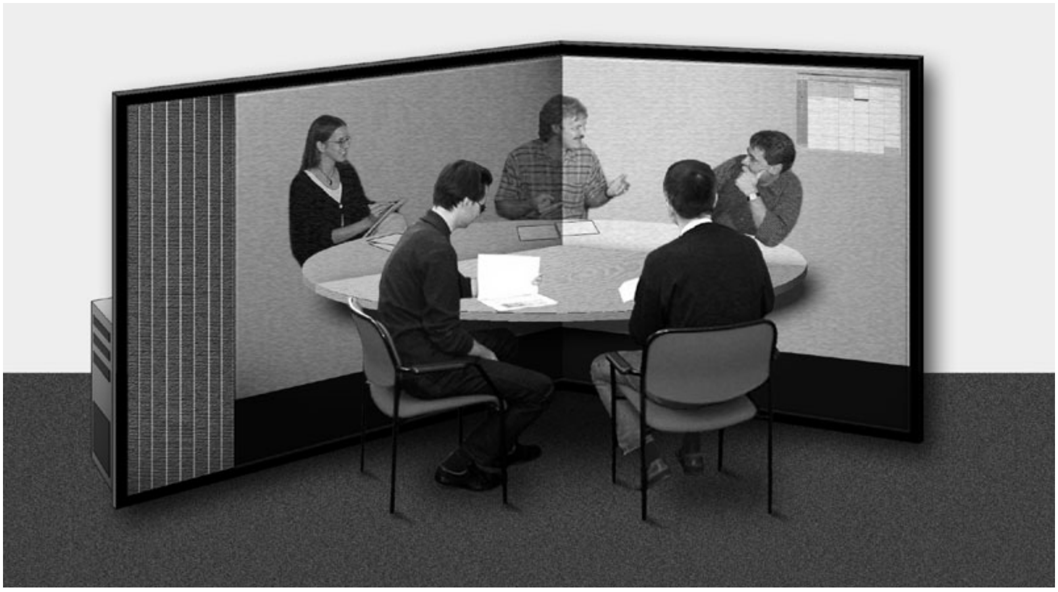
Live video sharing using external devices have become a day-to-day activity these days. From mobile devices to desktop webcam, people use different mediums to communicate. Skype [9], Periscope [10], FaceTime [11] and Facebook live [12] are some instances of remote collaboration that take place on a regular basis. The usage of wearable devices is growing rapidly around the world for telecommunication and remote collaboration [8]. My research extends on earlier research on creating shared emotional experiences [13], Affective Computing [14] and user theories explored in gaze transfer in remote collaboration [15], real-time mutual gaze perception [16] and coherent spatial referencing using shared gaze [5]. This section would expand on these concepts and bring out the research outcomes that would be potentially covered in my thesis.

2.1 Video conferencing and collaborative wearable systems

Recent research on wearable technologies advise its widespread benefit for field workers on remote collaboration and task coordination. CMU researchers found out that remote collaborative systems had a positive impact on maintenance duties on bicycles and aircrafts [17],[18]. Researchers from the University of Washington proposed a prototype which concluded a positive impact on 3D immersive collaborative environment through the usage of wearable technologies [19], [20].

Taking a step forward from this would be immersive 3D video conferencing and gaze tracking which allows the users to have a natural flow of conversations. This elevates the standards of human-centred communication. Researchers in video conferencing have come up with solutions such as an immersive tele-presence [21] using a shared virtual table, assuring direct eye contact and gesture representation [see Figure 2.1].

GAZE-2 [22] is another novel advancement in video conferencing which uses eye-controlled camera direction that establishes a parallax-free gaze connection. GAZE-2 automates the video channel using an eye tracker to avoid parallax errors. One of the several findings of the GAZE-2 researchers is that the camera movements do not disturb the gaze perception.



[Figure 2.1] Objective of Immersive Video conference [21]

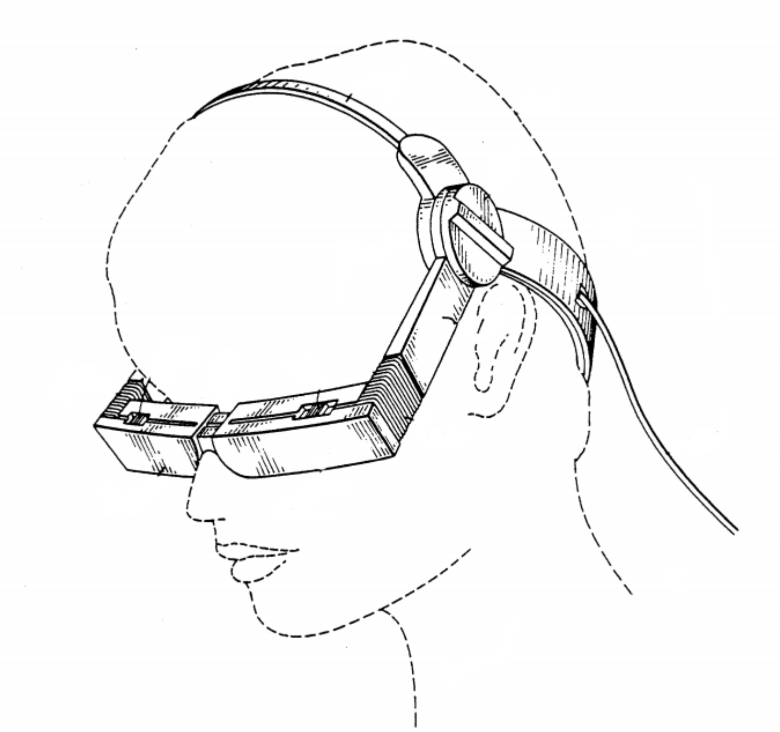
One of the main constraints of video communication systems is that the participants do not feel as connected as in a live environment. Modern researchers however have been exploring prototypes that could make them feel a lot more closer even in a video conferencing system. Neustaedter et al. [23] suggested that the communication process becomes more 'live' when the participants are aware of the location, weather, activities and status (eg. Healthiness). Inkpen's research [24] reinforced Neustaedter's study through his development of a system wherein friends and families shared an external experience. Yarosh et al. [25] further extended this research by studying the challenges of kids playing through remote collaboration focussing on their responsiveness, prominence and intersubjectivity. Similarly, Scwartz et al. [24] tested the value of external cues in a virtual environment by using one mouse and two mice based instructions while children playing a collaborative game [see Figure 2.2].

[Figure 2.2] Kids playing on a collaborative environment guiding actions to other kids through 1. Paper 2. One mouse 3. Two mice [26]

Telecommunication devices and services are the primary technologies designed to aid collaborative work. Video conferencing is considered as an advanced system that bridges telephonic communication and face-to-face communication. Being available for several decades, video conferencing has been proved as the one of the most favourable technology for remote collaboration [27]. Long distance collaboration avoiding the hassles of travelling and increasing flexibility in communication is possible through video conferencing. Even though the primary focus on the introduction of video conferencing systems were on white-collar workers in corporate environment, these systems are now available even at the household level where friends/families could interact with each other over long distance.

Wearable technologies on the other hand are intended as a stand-alone system supporting context-based communication of information but not at an interpersonal level of collaboration [see Figure 2.3]. Later, the technology has been proposed for use in remote collaboration with video conferencing for technical support and repair activities [28]. Kraut et al. [29] found that visual information to be one of the most useful resource in collaborative physical tasks. Fussell et al. [30] studied the effects and benefits of shared visual context on collaborative work .The results showed positive effects of using shared visual cues but had a few queries about the implementation of the shared visual system over the current video conferencing systems for implementation in the commercial market.

Armstrong et al. [31] prototyped an immersive environment that displayed a diabetic limb salvage surgery with the help of a Head-mounted display (Google Glass). This prototype helped display a live surgery to other surgical researchers with the use of Google Glass through Google hangouts. Other surgeons were able to communicate in the live environment and assist with the surgery through MRI images and clinical diagrams.



[Figure 2.3] Head-Mounted display for miniature video display system [32]

Wearable systems hence have become a preference in remote collaborative research due to its widespread positive impact with its convenience and ease-of-use. Head-mounted displays provide the convenience of immersing the user in a synthetically created surrounding. These units have varied applications to day-to-day users in the field of aviation, engineering, science, medicine, gaming, sports and even in television production. There are several commercial manufacturers of head-mounted displays including Canon [33], Oculus VR [34], Olympus [35]and Vuzix [36] to name a few.

2.2 Eye tracking and visual attention

Eye tracking [37] is one of the modern methodologies of research where a user's eye movements is tracked and recorded to identify where the gaze is directed in a specific portion of a visible field. Humans often switch attention to a particular portion (even for a short moment) of an object or location of interest. Using this tracking method researchers could get more insights into what captured the attention of the research participant and further provide hints on how the user apprehended the image they were looking at [38]. In my research, eye tracking was used to provide their partners an insight on where they were looking at in the remotely shared environment. The process is carried out with the help of an eye tracker which is an instrument to measure eye displacement and location. Although there are several methods to measure eye placements, the primary methods to carry out this process are:

- Tracking the direction of an object such as a contact lens attached to the eye
- Optical footprints of the eye captured without direct eye contact using computer vision
- Computation of electric ambiance through electrodes surfaced around the eyes

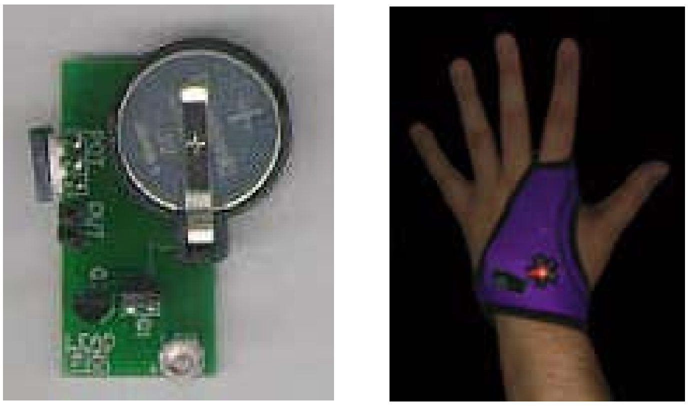
Some of the varied usage of eye tracking are in the field of system usability, training simulators, virtual reality, sports training and of course cognitive studies. In recent times, there have been various commercial applications of eye tracking to support the prevalent use of this system.

The use of eye tracking could be applied to its full potential when we understand where the user's attention is diverted to in a constantly differing visual environment and how their vision affects their actions and subsequently their partner's actions. Although there has been extensive research in similar fields such as the study by Henderson & Ferreira [26] on the interaction between vocal communication and visual perception, problem solving through the use of gaze movements by Reingold et al. [39], the impacts of visual perception in collaborative tasks by Bangerter and Clarke [40], the creation of an immersive environment for remote collaboration through gaze perception has still been a challenge.

2.3 Shared emotional experience and affective computing

Creating a remote collaborative environment where two or more humans could interact and share their emotional experience over long distance has gained attention with the widespread usage of video communication. The concept of shared emotional experience with a remote user aids in understanding and empathising from the perspective of the person sharing the experience. Humans express emotions to computer by nature but the computer does not typically pick up on those emotions during an interaction. Communication of emotions occurs only when the transferred emotion is received as well. Affective computing [41] concentrates on picking up deliberate emotional innuendo from humans to computers and enables machines to identify suggestive patterns of human expressions.

One such instance of Affective Computing hardware is the Galvactivator which is technically a hand glove that identifies and communicates skin conductivity [42]. The communication through skin reaction is known as electrodermal response where the skin on external or internal application of physiologically arousing stimuli becomes a better conductor of electricity [see Figure 2.4]. This arousal is one of the pointers to emotional activation and is picked up by the Galvactivator. This information is further stored as emotional responses as a part of affective computing.

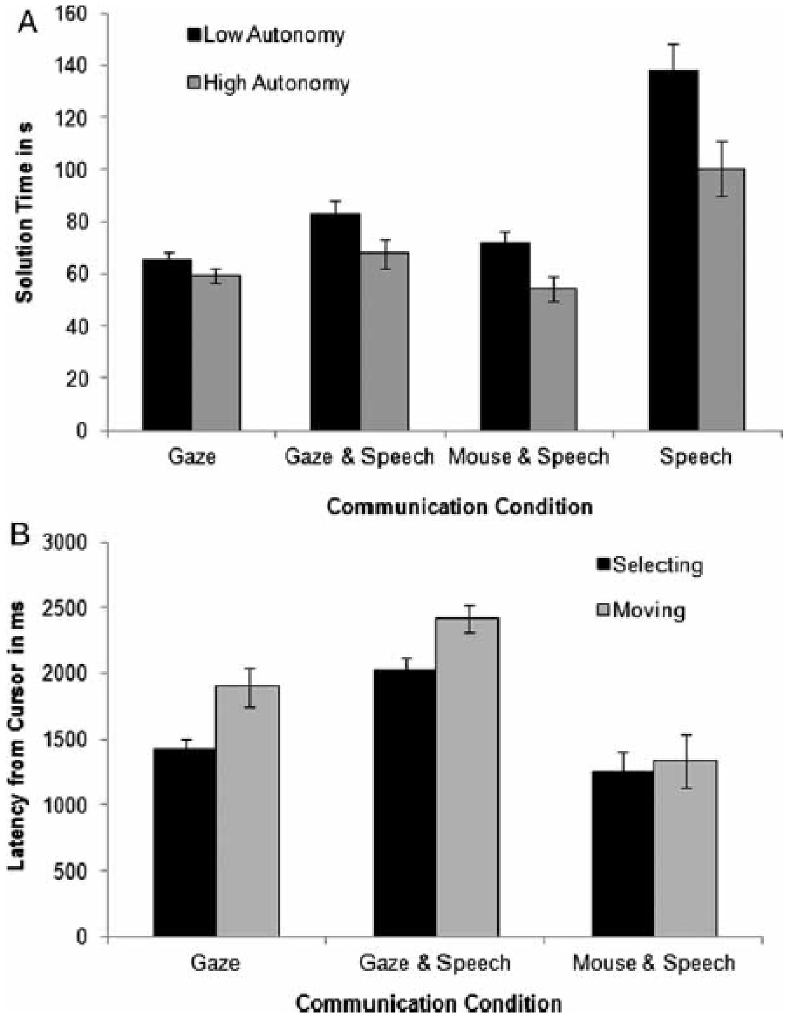


[Figure 2.4] Galvactivator circuit board (left) and the device on hand (right) [42]

Over the years, there have been many other instances of Affective computing that help the machine to identify human emotions and thus enabling an extended human-computer interaction. Some extensions in “affective wearables” include a Cybernetic wearable camera that identifies arousal response and saves user’s favourite videos [43] and a programmed “DJ” that analyses and monitors user’s mood and shuffles the song playlist based on the mood [44]. Several factors could be used for efficient Affective Computing such as speech, facial expression and in my experiment the focus would be on eye gaze.

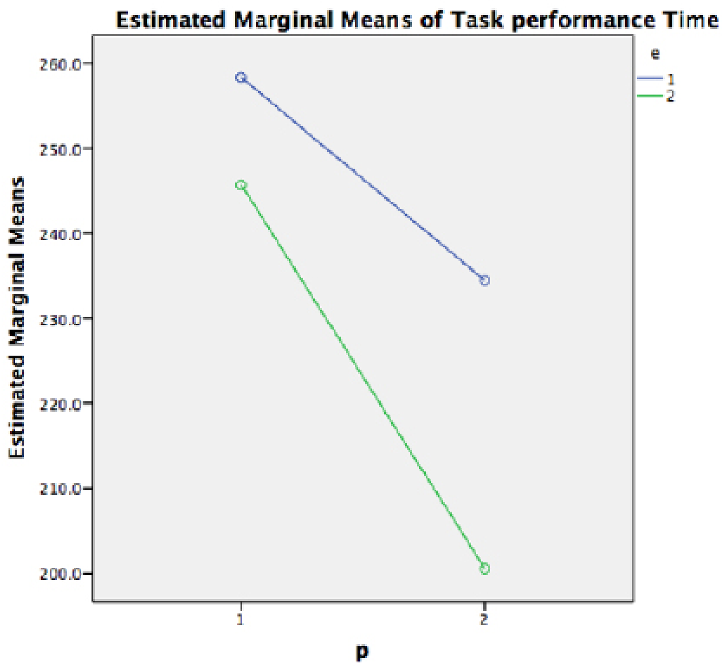
2.4 Shared gaze in remote collaboration

Research conducted by Muller et al. [15] showed that the participants received their instructions more clearly from their partners when gaze information was used in addition to speech. Their experiment required the participants to carry out a collaborative task in four different conditions (1) gaze, (2) voice, (3) gaze and voice, or (4) mouse and voice. The experiment was conducted in a desktop system by sharing a GUI screen using a puzzle application. The results proved that using the mouse or gaze cues with speech, the instructors were able to better communicate the instructions in performing the given tasks [see Figure 2.5]. However, using just the gaze information without speech were proved to be less effective in comparison to just speech. During the experiment, the users tried to dodge ambiguities by using vocal cues along with the mouse/gaze pointers that was shared with their partner. In conclusion, the results disregarded the use of only mouse/gaze pointers sans the voice in a collaborative task since the instructions seemed to be tentative and ambiguous, thus obscuring the process in their spatial referencing task. Extending on their research using a shared GUI screen, this thesis aims to research on the effectiveness of gaze tracking in a wearable video conferencing system.



[Figure 2.5] Solution time for experimental conditions (A) and latencies from first cursor reference on a piece to selecting and moving, depending on the type of cursor transfer (B) [15]

K. Gupta [8] designed an experiment to study the importance of wearable tools for affective remote collaboration, in which he studied the advancements of devices that recognize human feelings and approaches. This was a vice-versa platform of my research where he used an eye-tracker at the local-users end who were given instructions from the remote collaborator in conducting a collaborative task. He found that over 75 percent of the participants felt that the introduction of an additional input (eyetracker and mouse pointer) enhanced the communication process in easily guiding their commands towards the execution of the given task [see Figure 2.6]. From the qualitative feedback from his experiment, one of the participants at the local-users end mentioned that the eyetracker helped him pass the information to his remotely present partner on exactly where he has to look avoiding the stress in vocally communicating the exact location of the task element. He concluded that the time taken to complete the given task was significantly lesser using the eye tracker in comparison to regular vocal communication and using a mouse pointer. The errors recorded were also evidently lower with the introduction of gaze pointers. In contrast to Gupta's experiment this thesis would analyse the use of gaze cues from the remote user's end who would instruct the local users on a collaborative task platform.



[Figure 2.6] Task Time Performance: Interaction between conditions; voice (blue) and eye tracker (green) [8]

2.5 Summary

This chapter discussed video conferencing, head-mounted displays and eye tracking devices and reviewed previous research on the usage of these technologies in a remote collaborative system. From the past research, we can understand that there is ample scope for research on the benefits of gaze perception considering its vast usages as discussed earlier. Cherubini et al. [45] highlighted some of the potential uses of collaboration through eye gaze and its importance in virtual co-presence in their research. The limitations of using gaze cues in remote collaboration is that the viewers might get confused when making assumptions on where the gazer might be viewing in their environment unless it is supported by vocal cues from the remote user's end. This has been discussed in earlier research by Lobmeier, Fisher and Schwaninger [46].

My thesis intends to explore creating an effective collaborative environment with inspiration from the collective results from previous studies and using the different technologies in aiding the process. The focus is on the gaze sharing from the remote-user's end who is giving instructions to the local user, in ultimately finding the possible outcomes of different cues in performing a collaborative task. Based on the findings from the previous results, the conditions that were tested in my experiment are *voice, mouse + voice* and *gaze + voice*.

Through quantitative analysis and qualitative feedback, my research would provide an understanding of effects of the different cues in a task-based collaboration and highlight the varied potentials of these cues. Finally, the results would be analysed to provide a solution to the limitations of video conferencing as mentioned earlier and also discuss on opportunities for future work on similar research.

3

Design

My thesis research developed a system to enable immersive multimodal interaction using shared gaze for reciprocated attention. The two-way communication featured a local user who captures his/her vision using a wearable head mounted video capturing device which transmitted the captured video to a remote user (instructor) sitting in front of a monitor with an eye tracker, to watch the video and share their calibrated eye gaze. User experiments were conducted using different conditions to test if the display of a gaze tracker at the remote user's end would enhance the remote communication process.

3.1 Idea conceptualization

A design thinking process was used at different phases to understand the potential users and to ideate concepts and prototype design solutions. The design solutions were later tested with the pilot users. To understand the users I first had to understand the context and culture of the users that would be using the prototype. Through direct observation and qualitative data from previously conducted researchers it was easy to create personas for the target users. Some of the identified personas are discussed in the next section of this chapter. Through the demographic identification, I was able to empathize with the users and tackle the potential challenges that might take place during the process of experimentation.

After diagnosing potential problems through the pilot tests, it was simple to ideate and prototype without the constraints of the existing solutions. This section explains the concept generation process with the help of various research methodologies including brainstorming sessions, paper prototypes, wireframing and conducting focus group interviews before generating the finished prototype.

3.2 User research

User research was conducted at the beginning of the research phase where the users were interviewed about the efficiency of the system and how much of a difference this made in their interaction. Focus group interviews were conducted to identify problems faced in these interactions and potential key outcomes. A sample focus group of six students and staff from the university were interviewed to understand their usage of video conferencing systems in their day-to-day life. There were four male and two female participants in an informal setting where they were asked a few questions on the topic.

The findings from the focus group reported that although they were regular users of video conferencing systems, most of them mentioned that they had not used a video conferencing system to perform a collaborative task. One of them mentioned that the use of external cues would be helpful when assisting with technical support to their clients. To help them get a perspective of where this system might be effective, they were given a few scenarios. One of the scenarios of the collaborative system was to perform a plumbing task at their home with an expert plumber guiding them through the process from a remote collaborative environment. Most of the participants felt that the cues would have a lot of scope in a collaborative task when they were given a real-life scenario.

The following sections elaborate on the concepts and methods that I used to create and iterate the final design prototype. The pre-production phase of my design process was covered at this stage.

3.3 Users and use cases

For this study, it was important to understand the users and construct use cases to collect samples for the study of the defined hypothesis. There are various uses of remote collaborative systems and services to support group work, technical support, conduct interviews, and communicate with offshore/onsite clients, to give some examples. Due to its widespread relevance identifying the samples with prior experience for our collaborative task was simple.

Use cases were made to give a perspective of how this system would be used in a real-life scenario. These use cases helped the participants in the user experiment (described in Chapter 5) to empathize with the scenarios where this system might be effective. Some of the use-cases identified were as follows:

- A novice onsite plumber who is not sure about a few fixes in a house and so collaborates with an expert remotely to guide him.
- A construction worker who wants to present an issue to their supervisor in the middle of work but does not want to have to wait until their supervisor get to the location before getting approval.
- Firemen are remotely guided by the control room to exactly where maximum damage has occurred in a multi-story building.
- A novice pilot learns to understand the buttons in the simulated cockpit from an experienced pilot communicating with him remotely.

3.4 Brainstorming sessions and story boarding

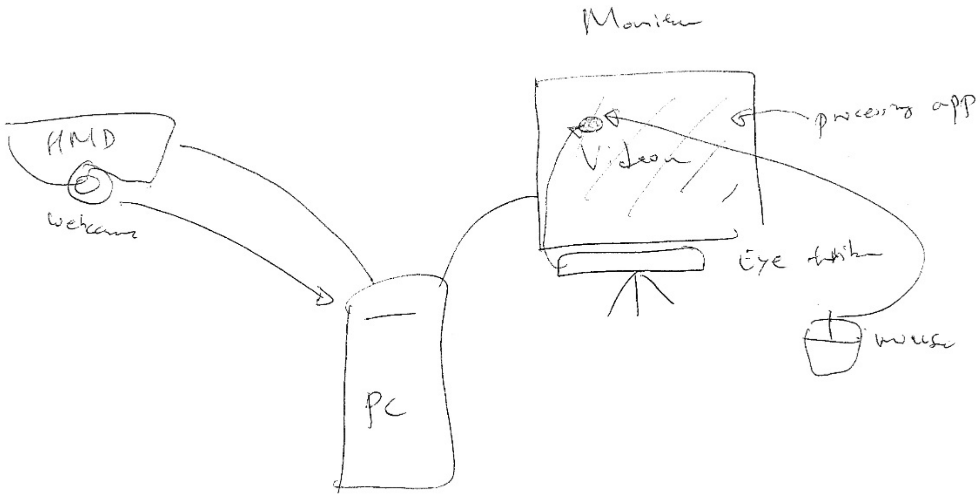
Several brainstorming sessions were conducted before the prototyping process. With insights from previous researchers on similar topics and the guidance of my supervisor, we were able to identify a few assumptions and potential challenges that might occur during the user evaluation process. Some of the assumptions from the brainstorming sessions as follows:

- All the information displayed to the remote user would be from the perspective of the local user wearing the HMD. The remote user would instruct the local user to move around the projected prototype to carry out the collaborative task.
- The remote user would have their eye calibrated to the eye tracker in front of them. This information would be overlaid on the local user's camera feed on the Head-mounted device to guide them to carry out the instructed task. The same applies to the mouse pointer.
- In order to simulate a real-life scenario, the remote user/instructor might have to be placed away from the local user's vision or have a blocker placed before them to focus on the camera feed displayed on the desktop.

With these assumptions in mind and with the joint efforts of other researchers from the HIT Lab NZ, I was able to conceptualize various ideas before testing the prototypes.

3.5 Prototype system design

Based on the brainstorming sessions, important aspects of the prototype were decided. Two final ideas were shortlisted from a list of concepts. The ideas were further sketched out and a design workflow was made to understand the process of the collaborative task [see Figure 3.1].

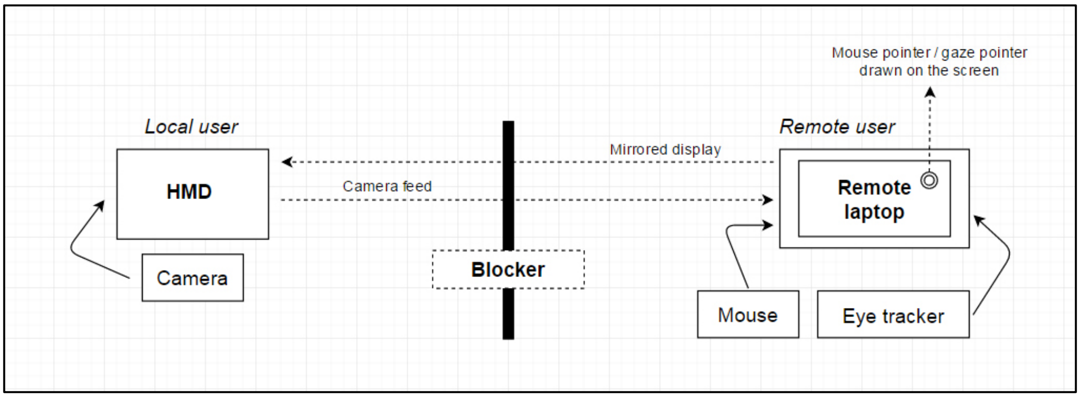


[Figure 3.1] First draft sketch of the prototype design

With the creation of use cases, we were also able to understand the limitations in hardware and the physical requirements to be followed for the smooth experience using the prototype:

- Ease-of-use is a necessity if used on a day-to-day basis. Considering that the local-user would have a wearable device, the HMD must be lightweight and convenient to use.
- The HMD must be able to provide maximum coverage of what's in front of the user.
- Visuals cues must be clear and understandable.
- The process of the collaboration should be simple and comprehensible.

The concept that was chosen after ideating and identifying the potential process was to use a remote laptop connected to a head-mounted display for the remote users and local users respectively. A camera was decided to be connected to the HMD to receive and feed the visuals before the local users to the remote user's laptop. For testing out the different conditions, a mouse and an eye tracker were chosen to be connected to the remote user's laptop. It was also decided that the camera feed display on the laptop would be mirrored onto the HMD so that both users look at the same screen. Figure 3.2 shows a system diagram of the prototype system.



[Figure 3.2] System diagram of the prototype system design

In order to replicate a real-life scenario of having the remote users distanced from the local users a blocker was decided to be placed between the participants. This way the remote users would be able to see the local user's visuals only from their laptop screen. A bold translucent mouse pointer would appear when the remote users participate in the collaboration process using their mouse. Similarly, a pointer would appear when they use their calibrated eye gaze while testing the gaze condition.

The main concern while representing the collaborative prototype was how the HMD performed in presenting the information before them along with the visual cues passed on from the remote collaborator. The cues have to be simple, and yet they should not affect the real-time visuals before them. In an ideal world, the local user should have the ability to easily perform the collaborative task with minimum errors, and should not be distracted from the external cues from the remote collaborator.

3.6 Design of the experimental task and environment

After constructing a solid design system, the next step was to design the experimental task and environment. The experimental task plays a key role in the user evaluation process since the task had to resemble a real-life usage scenario while the participants collaborate on the remote environment. A few ideas were brainstormed for this purpose and two ideas were finally shortlisted for the final experimental task.

The first idea for the user experiment was to use a puzzle/LEGO setup with both the participants facing away from each other. The instructor/remote user would be given the final sketch of the puzzle/LEGO figure to guide the local user using the three conditions (voice, mouse + voice, gaze + voice). The second idea was to simulate a control panel with a set of different action buttons and indicative icons next to them. The remote users would be provided with a list of actions that they would instruct the local users to perform on the control panel. This was inspired by a few simulation VR applications that are available to train participants on the usage of the application before interacting with the live environment. One example is the CEFA Cockpit Emulator [47] and Pedagogical Flight Simulator for longitudinal airplane flight [see Figure 3.3].

[Figure 3.3] CEFA Aviation Cockpit emulator for Flight Analysis [47]

The conceptualised ideas on the design concept were presented to a few researchers for feedback and analysis. Some of the outcomes from the feedback are explained below:

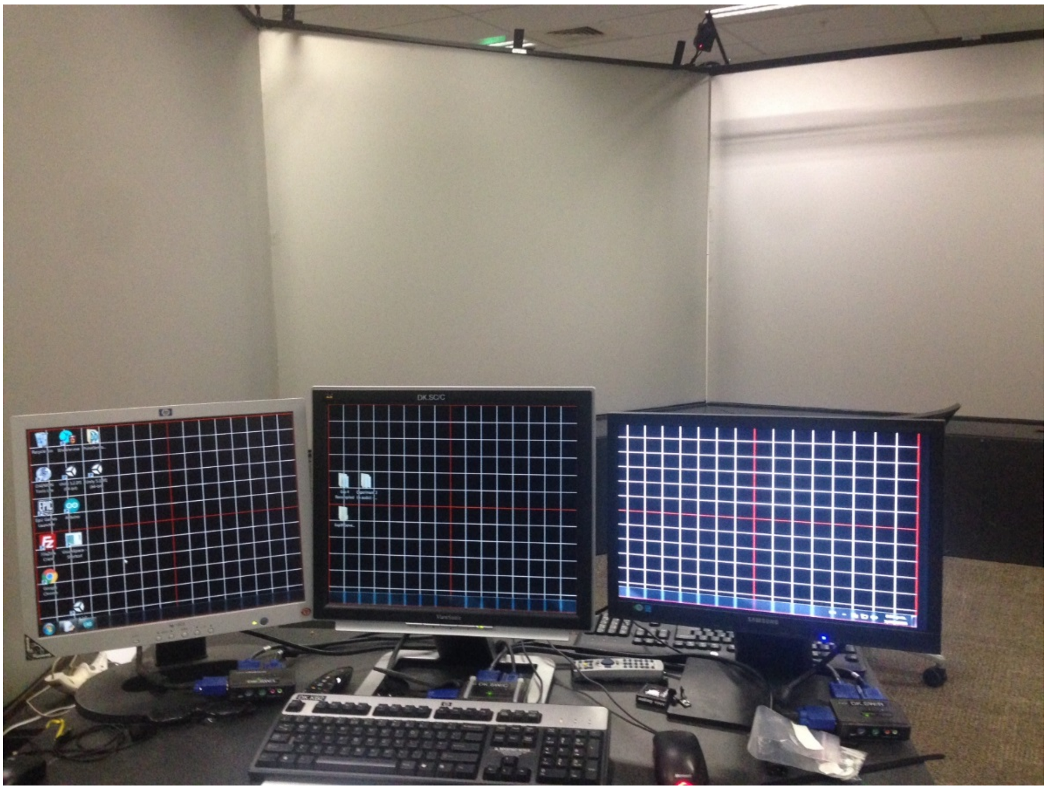
- Similar tasks would be carried out by the participants for every condition of the research.
- Design elements would be varied within a range of similar patterns to make the identification process challenging to the remote users.
- Task completion time and measured using the different conditions would help understand the most effective condition.
- Instructions would be given by the remote user through speech even if an additional condition is implemented during the research process.
- The interface would be on a projected display while the local user would have to move around in order to present the information across the interface.

After considering the outcomes from the brainstorming sessions and the feedback on the experimental task, the second option was opted for the final user experiment. The reason for choosing this idea was that the local users had space to move around and interact with the prototype while the remote users instruct them from a distant desktop. This simulates a real environment where people might use the collaborative setting.

In conclusion, we decided that the prototype would be a large-screen display of various interactions (buttons, sliders, toggle buttons, knobs, etc.) that would be presented in front of the local user. This would be fed to the instructor's desktop display through a camera placed before the local user's HMD. The instructors would be given a set of instructions that they would guide the local users in carrying out. They would also judge whether the action is carried out before moving from one action to the next.

The finalised user experiment consisted of the following:

- A projected display consisting of icons next to different action buttons including press buttons, toggle buttons, switches, knobs and sliders
- Local user's wearing a HMD with a mirrored display of the remote user's desktop to carry out the collaborative task as instructed.
- Remote desktop for the instructor where they would be able to see a part of the projected display through the camera feed from the local user
- A set of action slides that would be assigned to the instructor
- A timer to calculate the time taken to complete the assigned set of actions for every condition



[Figure 3.4] Multiple monitor setup at HIT Lab NZ Vision Space with the Vision Space wall in the background

The best type of multi-utility control for presenting simulations is made of tiled demonstration structures using large projection display. We used the Vision Space [see Figure 3.4] facility at the lab, which is an immersive back-projection display setup with three screens. Each screen has a size of 2.45m by 1.8m, and projector with a resolution of 1024 x 768 pixels.

4

Prototype Implementation

On completion of the design concept with the help of the brainstorming sessions and paper sketches, the next phase was the implementation of a working prototype constructed based on the designs. This section elaborates on my thought process using the various hardware and software apparatuses used in building the prototype. The finalized prototype was later utilized in conducting user assessment (discussed in detail in the next chapter).

4.1 Hardware

This section briefs on the technology used for the testing and evaluation phase of my research topic.

4.1.1 VUZIX STAR 1200XLD

A part of the Vuzix family of head-mounted displays, the STAR 1200XLD [36] is a photosensitive transparent AR device that provisions stereo sound and 2D or 3D visuals to practically all HDMI-out systems [see Figure 4.1]. This solo eyewear device allows connections to an array of products including desktop computers, laptops and mobile interfaces such as tablets, smartphones or even a 3D Blu-ray system. This HMD device has the ability to present a 75 inch simulated display when viewed from a distance of 10 feet (approx. 3m) and displays 2D or 3D video in high resolution digital clarity. Adaptable eye-separation and frames that tilt to adjust to the comfort of the user are the main features of this portable device. A HD PRO Webcam C90 [48] was attached using Velcro onto the local user's HMD to display the camera feed to the remote user's desktop. This camera allows a true-to-life 1080p video capture through a H.264 compression technology.

[Figure 4.1] Studio shot of the Vuzix STAR 1200XLD (Credits: Vuzix official website) [36]

One of the most distracting aspects of using the HMD is the cause of motion-sickness to the users (as is the case with most of the VR displays). The main reasoning behind the motion-sickness is the struggle between the figures displayed to the optical senses that are moving in a VR display, while the vestibular senses inform the user's brain that they are in a static location. However, we concluded that the users would not be affected much by this since the information displayed on the HMD is just the camera feed from what is before them. Hence, it would just be a case of displaying real-time information with additional cues for the different testing conditions.

4.1.2 EYETRIBE

The Eye Tribe [49] is one of the recent commercially available eye tracking products that facilitates gaze control over mobile and desktop devices [see Figure 4.2]. It supports hands-free navigation across webpages and web applications with added features such as gaze stimulated login, superior gaming involvement and cloud based user assignation analytics. Measuring 20 x 1.9 x 1.9 cm, the Eye Tribe has a spatial resolution of 0.1° (RMS) with 30 Hz and 60 Hz mode sampling rates.

[Figure 4.2] Studio shot of the EyeTribe eye tracker (credits: EyeTribe official website) [49]

In my experiment, the Eye Tribe was calibrated at the remote users end with a 12 point calibration process where the users move their gaze towards the pointers on the desktop. This process was conducted a few times to ensure that the tracker is aligned to the respective user's gaze and they are comfortable using the tracking system.

4.1.3 LAPTOP

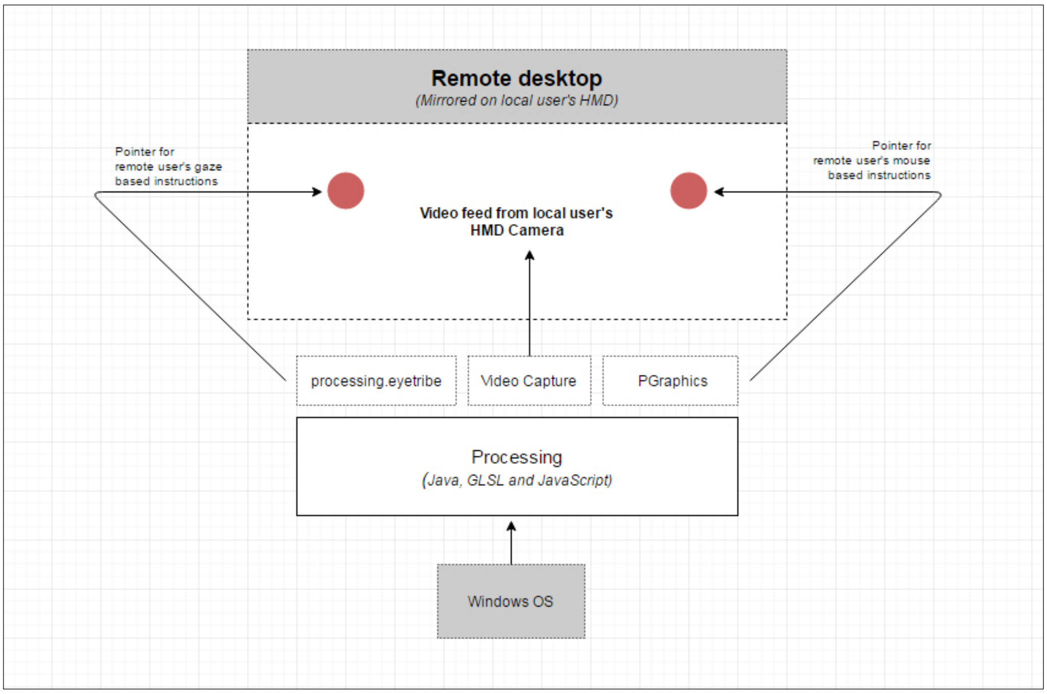
A Toshiba Satellite P70-b [50] was used at the remote user's end for the user experiment [see Figure 4.3]. This system comprises of an Intel(R) Core i7-4720HQ CPU with a 2.60 GHz Processor. With a 16GB RAM, this 64-bit Operating System runs on Windows 10. With a widescreen 17.3 inch 16:9, 1920x1080 pixel display, the laptop provided a wide display of the camera feed at the instructors end.

[Figure 4.3] Studio shot of the Toshiba Satellite p70-b (Credits to Toshiba Official Website) [50]

4.2 SOFTWARE

The software used for the implementation of my prototype was Processing [51]. Processing is an open source programming language and Integrated development environment (IDE) built for basic programming in a cross-platform environment. The software program for my experiment was developed using Java, GLSL and JavaScript. The 'Video Capture' library was installed to receive the camera feed from the HMD and 'PGraphics' function was used to draw the cues from the mouse pointer and gaze tracker.

The camera feed displayed on the remote user's desktop was duplicated to the local user's HMD so that both the participants were seeing the same interface [see Figure 4.4].



[Figure 4.4] Block diagram of the working of the software components

The purpose of the application developed is to test the three different conditions in the remote collaborative environment. To ensure that the setup represents a real-world scenario, there were a few necessities that were kept in mind while developing the application:

- The local-user must be able to easily comprehend the cues from the remote user
- Remote-users should have the ease-of-usage while providing instructions to their partners
- Both users must stay focused within the developed environment
- Remote-users must pick-up the entire camera-feed from the local-users HMD camera

[Figure 4.5] A view of the remote user's desktop screen with the camera feed from the local user's HMD camera and the mouse pointer (red dot)

Figure 4.5 shows a preview of the remote user's desktop where the local user is interacting with the projected control panel after receiving the respective instructions from the remote user. This screen was also mirrored onto the local user's HMD during the remote collaborative process.

4.3 SUMMARY

The development of the high-fidelity prototype was discussed in detail in this chapter along a summary of hardware technologies used in the prototype system [see Figure 4.6]. The reasoning behind choosing the used hardware was also explained. The main elements involved in building the prototype software was also elaborated in this section. Java and Javascript was the main formats used in Processing software in the development of the prototype application. A strong design thinking process was implemented to mimic a real-life scenario for the experiment. The following chapters expands on the evaluation of the prototype system.

[Figure 4.6] The experimental setup from the remote users end with the control panel projected in the Vision Space Display

5

Evaluation

The user study led to compare performances of the three conditions in my research experiment is elaborated in this chapter. The aim of the user study and design is briefed in the first part of the chapter followed by the explanation on the study results. Results of the experiment and further analysis is presented at the end. Consent forms and information papers could be found in Appendix A.

5.1 EVALUATION PURPOSE

The purpose of the evaluation was to study and understand the effects of the three different conditions that were tested using the high-fidelity prototype. A user experiment was conducted to compare the different conditions and attain the results. A set of qualitative feedback and quantitative measurements based on task performance time and errors were collected for the purpose of the evaluation.

5.2 EXPERIMENT DESIGN

To study the different cues in the remote collaborative environment using the HMD and gaze tracker, a within-subject study was performed. The experimental conditions of the experiment are:

- Voice cues
- Mouse + voice cues
- Gaze + voice cues

The dependent variables in my experiment are:

- Task time completion to complete the various actions
- Number of errors made while performing the actions
- Overall system usability
- Social presence
- Ranking based on user’s preference

The overall effect of each of the conditions on user experience was measured using qualitative feedback at the end of each condition. The participants were requested to keep the overall system usability and social presence in mind while answering the feedback questionnaire. This way they had more perspective on how effective each of the conditions were in completing their assigned task with the help of the local users.

5.2.1 HYPOTHESIS

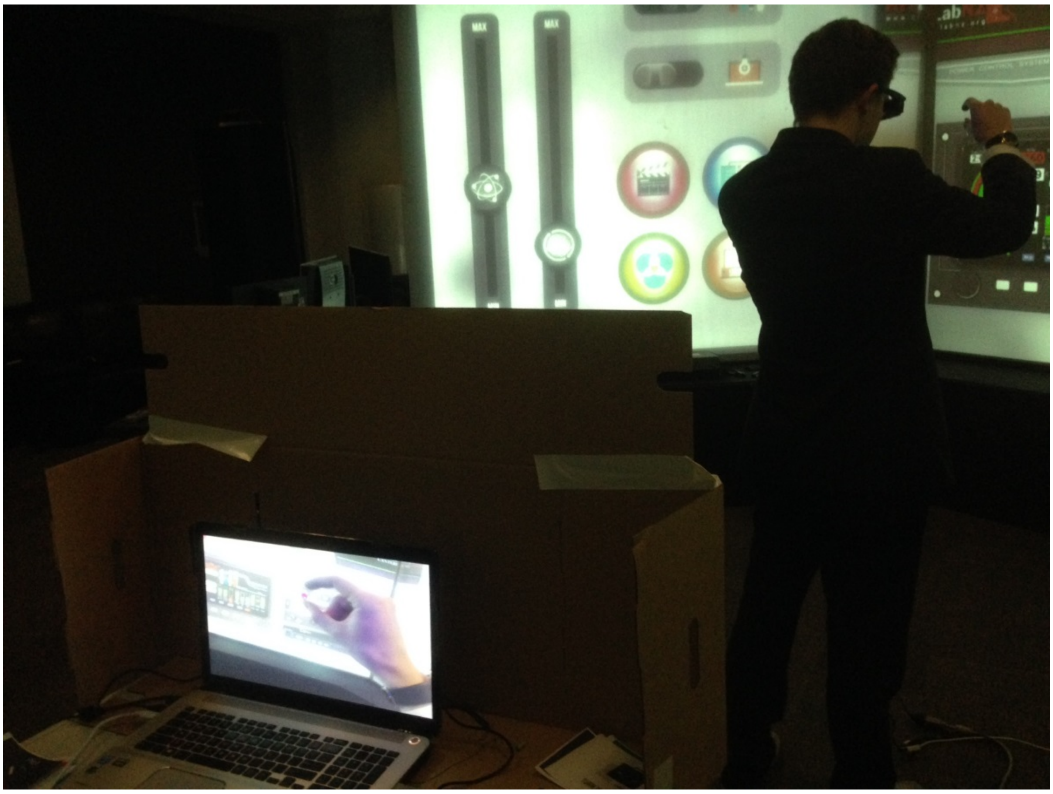
Considering that the gaze tracker provides an ease-of-use and communication through direct observation in comparison to the manually inputted information using mouse pointers and time consuming vocal cues, Hypothesis 2 was formulated. Hypothesis 3 was formulated with the knowledge that people are usually immersed within the collaboration process however the efficiency of the collaborative process might change with changes on the emotional level of interaction. Hypothesis 1 and 4 was arrived with the knowledge gained from related research on the effectiveness of gaze cues.

1. There is significant difference in the task completion time and the participants are able to complete the task faster using gaze + voice cues.
2. There is significant difference in the usability of the prototype when the remote collaborator/instructor guides the local user wearing the Head-mounted device. The usability is in favour of eye gaze over mouse pointer and voice cues.
3. There is significant difference on a social presence level affecting the users among the three conditions while using the prototype.
4. Gaze + voice cues is the most preferred condition in overall ranking among the three test conditions in carrying out a remote collaborative task.

5.2.2 EXPERIMENTAL SETUP

The user study took place at the Vision Space in HIT Lab NZ and the materials used for the study were:

- A Laptop
- Vuzix STAR 1200XLD head-mounted display
- Logitech HD PRO C90 web camera
- Eye Tribe eye tracker
- Vision Space desktop and wall for the prototype display
- Chairs and desks for setting up the materials
- Board to block the remote users view



[Figure 5.1] Remote user guiding the local user to turn a knob using his calibrated eye tracker

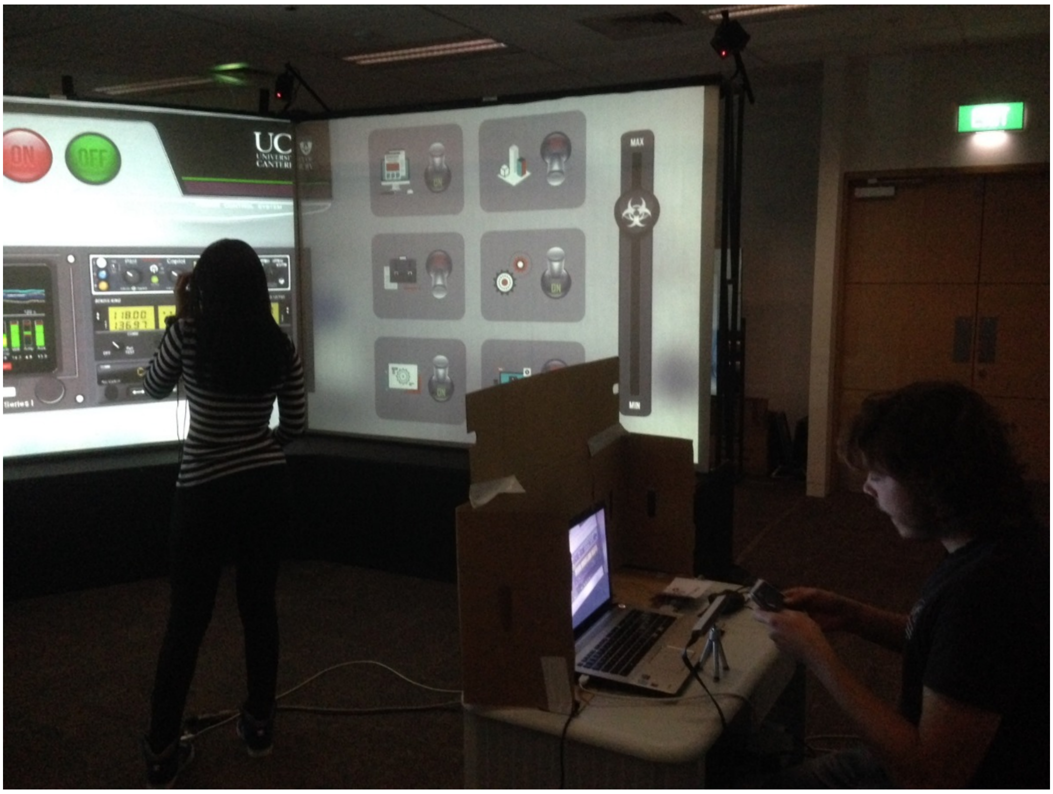
Figure 5.1 shows one of the user experiments where the local user is trying to turn a knob after being instructed by the remote user's eye gaze and voice guidance. As mentioned earlier, the projected control panel was made of different action buttons including press buttons, knobs, sliders and switches. Icons of different categories were placed next to the action buttons which were identified collaboratively based on the instruction sheets given to the remote users. Figure 5.2 shows an overview of the different action buttons that the participants identified and interacted during the user experiment.

[Figure 5.2] An overview of the different action buttons that were projected on the Vision Space wall.

The local-users with the attached Head-mounted device were be able to view a part of the Vision space screen displayed before them and the video from the camera attached on the HMD was transferred to the remote user's desktop. Since, the control panel for the experimental task was shown across three projection screens, the local-users had to move around the projected screen in order to carry out the assigned tasks that the instructor/remote user guides them to perform. With the help of the portrayed visuals on the screen and the guidance of the remote users, the local users were to follow instructions from the remote users and mimic different actions (such as pressing buttons, moving sliders from minimum to maximum, toggle buttons and turn off/on switches) on the projected control panel.

A table with the cardboard blocker was placed around the remote user's desk so that the remote users were not able to see the projected interfaces that were viewed by the local user. A laptop was placed on the remote user's desk from which they could view the camera feed from camera mounted on the local user's HMD. A mouse was connected to the laptop and used for the second condition and the Eye Tribe eye tracker was also connected for the third condition. Another desk was placed in the corner accompanied by chairs to be used by the users while filling out the questionnaires after each condition. The Vuzix HMD was connected to the laptop with a video cable and long USB extensions that allowed the local user's to move around the Vision Space wall to carry out the experimental tasks. The cables were held together with tape to avoid unnecessary mishaps during the process of the experiment

Figure 5.3 shows an overview of the entire experimental setup where the remote user is guiding the local user using his eye gaze (using Eye Tribe in front of him) to perform the assigned instruction. Details of the prototype system used in the experiment can be found in Chapter 4.



[Figure 5.3] Participant from previous research wearing the HMD and Eye Tracking Application running in the monitor behind

There were a few challenges that had to be tackled for the smooth running of the interface. For my experiment, the visual cues for the mouse pointer and the gaze pointer were both designed to look similar to provide clear guidance. The other challenge was to decide how much information should be available to the local user while testing the different conditions. We decided to keep it simple with the graphics so that the local user's display is not crowded with information.

5.2.3 EXPERIMENTAL TASK

The participants were required to complete a set of tasks using the three conditions within the best of their capability. The task for the local user was to follow the instruction given by the remote user, and the remote user was handed a series of instructions printed in A5 size papers as shown in Figure 5.4. Please see Appendix A for more examples of the instructions. These instructions were randomized and the order of the conditions were changed between different sets of participants. The order of the conditions were decided using balanced Latin Square design to counter balance the ordering effect.

The participants were informed that the tasks would be timed from when they began to perform the first instruction. The remote user had to ask their partner to move around the wall to identify the icons that pertained to one particular action. When they recognized the icon, they then instructed the local user to carry out the respective task assigned to that icon. The instructions given for the local users were toggle buttons, turn on/off switches, and move sliders to minimum/maximum, turn knobs and press buttons. The local users were asked to act out what was instructed, though the mock-up controls on the projection screen was not interactive.

[Figure 5.4] A sample of the action sheet that was handed over to the remote users

5.2.4 PROCEDURE

At the beginning of the experiment, the prototype system and process was explained to the participants and they were requested to read the information sheet. Once the participants were clear of the process, they were further requested to sign the consent form. They then filled-out a pre-experiment questionnaire that collected demographic information. This was collected to understand the participant's demographics for further analysis. The researcher then explained to them about the hardware that they would be using for the collaborative process before running a few practice trials. (The verbal script of the experiment is available in the Appendix A).

The remote users were seated in front of the laptop and then handed over the instruction sheets that they would be guiding their partners to perform actions on the displayed control panel. The local users were standing at the centre of the Vision Space wall. Their Head-mounted display was adjusted to fit their vision and they were asked to move around to make sure all panels of the wall is accessible to perform the actions instructed by the remote user. Once the participants were comfortable with the interface, the lights were dimmed down to ensure maximum visibility of the projected interface.

In each condition, participants had to complete five instructions. After completing each condition, the participants were taken to another desk and were requested to fill out questionnaires asking on usability and social presence level of the condition. No suggestions were given on the feedback to ensure there was no bias in their answers. Once the feedback form was finished the participants moved on to the next condition. The same was repeated for the other two conditions.

In the case of the gaze + voice condition, the remote users were placed before the Eye Tribe eye tracker and calibrated the tracker with a 12 point calibration panel using the Eye Tribe GUI. This was repeated a few times to make sure that the calibration was perfect before they moved on to perform their actions. Practice trials were given at the beginning of each condition so that the participants were at ease of using the interface.

At the end of the experiment, the participants were requested to fill out a post-experiment survey by ranking each condition and providing feedback on suggested improvement of the interface. This was followed by an unstructured debriefing interview in which the participants who were interested in more details about the experiment were told about my hypothesis and the potential outcomes that the researcher envisions from the background research.

5.2.5 MEASUREMENTS

A set of qualitative and quantitative data was collected from the remote users during the process of the experiment. The time taken for task completion for every condition was recorded for all participants. Demographic information including age and gender, how frequently the users were accustomed using a video collaborative system and their relationship with their experiment partner was covered in the pre-experiment questionnaire.

The usability of the interface was measured using the System Usability Scale [52] that the remote users had to answer at the end of each condition. Feedback on their collaborative experience was collected using the Social Presence survey [53] which included questions about their co-presence, attention allocation, perceived message understanding and perceived affective understanding of their partners while involved in the remote collaborative process. Perceived emotional interdependence and perceived behavioural interdependence had been left out from the original questionnaire since they weren't relevant to this study. The users had to write their collaborative feedback at the end of each condition as well. A copy of the questionnaire has been included in the Appendix A.

The clarity of the questions was ensured using a pilot study and the remote users were also guided in any case of any ambiguity in the survey form. All the survey questions had a 5 point Likert scale ranging from **1**: Strongly disagree to **5**: Strongly agree.

This was later compared using different tests:

- **Friedman test** to test the differences between the ordinal dependant variables
- **Wilcoxin signed-rank test** to compare each of the within-subject dependant variables when the results were ordinal or not normally distributed

5.2.6 PARTICIPANTS

Initial pilot tests were run with three sets of participants. The results from the pilot test did not contribute towards the final results of the study. The purpose of the pilot tests were just to identify early bugs and ensure smooth running of the final user experiment. Peers and supervisors were chosen for the pilot test to give me feedback on their experience and also improve the feedback questionnaire that would be solicited to the final participants.

Researchers from the lab who had not been involved in my brainstorming, responders to the posters that were circulated around the university and a few other associates were chosen to be the sample for my main experiment. 30 participants (15 pairs) were recruited for the user experiment. The tested participants were an almost equal mixture of men (55%) and women (45%) with their age brackets of 24-30. Almost 95 percent of them used video communication systems at least a few times a week and the remaining used it at least a few times a month. 64 percent of the participants knew their partners before as friends, 18 percent of them were related and the remaining were acquaintances.

5.3 RESULTS

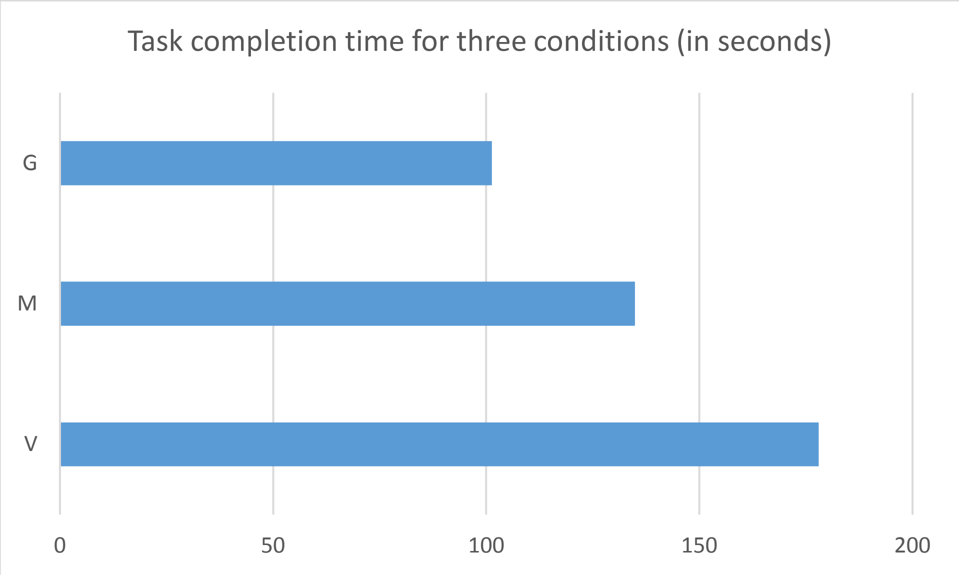
The former portion of this section elaborates on the quantitative data collected through the user experiment. The later briefs on the qualitative feedback collected from the post-experiment questionnaire at the end of the experiment. Some valuable feedback from the participants that was received through a debriefing interview is also included at the end of this section. The conditions in this section have been abbreviated to voice = V, mouse + voice = M, gaze + voice = G in the bar charts and box plots in the following sections.

5.3.1 TASK COMPLETION TIME

The task completion time (measured in seconds) for each of the condition for the 15 user experiments was recorded. After completing a normality test using Shapiro-Wilk test, the data was found to be normally distributed for voice ($W = .970$, $p = .864$), mouse + voice ($W = .972$, $p = .880$) and gaze + voice ($W = .950$, $p = .526$).

There was a statistically significant difference between conditions voice, mouse + voice and gaze + voice determined by a repeated measures one-way ANOVA ($F(2,28) =$

17.410, $p < .001$). A post-hoc Bonferroni test revealed that there was statistically significant difference between all three conditions, voice-mouse + voice ($p = .044$), voice-gaze + voice ($p < .001$) and mouse + voice – gaze + voice ($p = .040$).



[Figure 5.5] Mean of the task completion time for the three tested conditions (in seconds)

Overall results [see Figure 5.5] show that the participants were able to complete the task faster using gaze + voice cues (M = 101.34 sec, S.D. = 22.18) over mouse + voice (M = 134.8 sec, S.D. = 38.5) and voice (M = 178 sec, S.D. = 47.9 sec). Analysis of the results from the questionnaires that were answered at the end of each condition during the experiment have been elaborated in the following sections.

5.3.2 SYSTEM USABILITY SCALE

The quantitative data for the SUS was recorded through a set of ten questions at the end of each condition for all 15 participants who played the role of remote users. The questions were alternated between positive and negative queries.

Frequency of system usage, ease-of-use, convenience, consistency, integration, were some of the questions that the participants had to answer in this survey. The questionnaire was answered with a Likert-scale rating ranging from 1 (strongly disagree) to 5 (strongly agree) and then aggregated into a scale ranging from 0 to 100 by subtracting 1 from the odd numbered responses and subtracting 5 from the even numbered responses. This converts the scale from 0 to 4 (4 being the most positive response). Then, the total of the converted responses were multiplied by 2.5 which converts the range from 0 to 100 instead of 0 t 40.

The results were tested using the Friedman test for the within-subject design. A post-hoc test was carried out using the Wilcoxon-Signed Rank Test for comparing the pairs of conditions (alpha = 0.0167, Bonferroni correction applied). The three conditions have been abbreviated as V for Voice, M for mouse + voice and G for eye gaze + voice. Mean, Standard deviation, min-max scores and percentiles were calculated for the descriptive statistics of SUS [see Table 5.1].

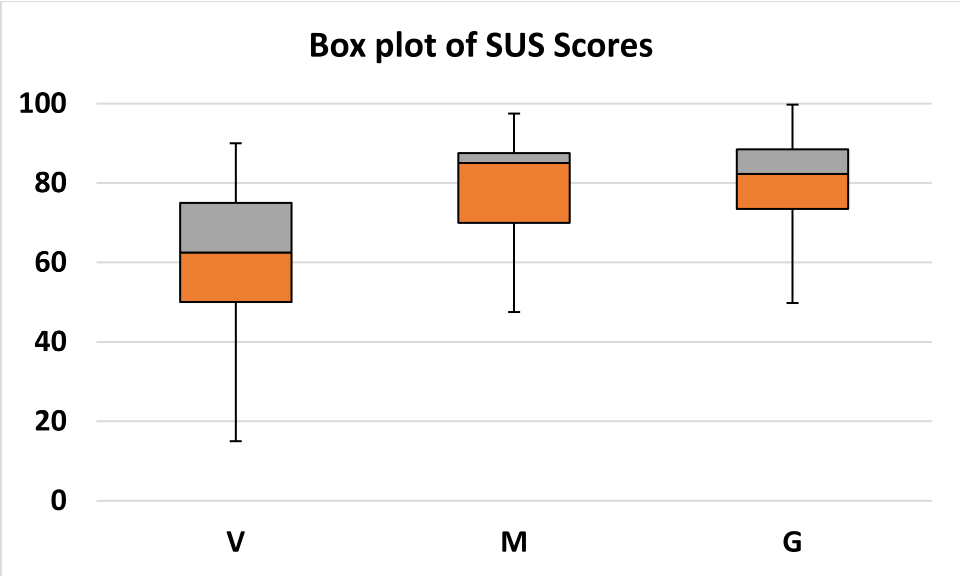
						Percentiles		
	N	Mean	Std. Deviation	Minimum	Maximum	25th	50th (Median)	75th
V	15	61.00	19.95	15.0	90.0	45.00	62.50	75.00
M	15	78.33	15.31	47.5	97.5	67.50	85.00	87.50
G	15	81.16	13.85	50.0	100.0	72.50	82.50	90.00

[Table 5.1] Descriptive Statistics of results from System Usability Scale

The mean SUS scores for V, M, and G were 61.0, 78.3, and 81.2 respectively. There was a statistically significant difference in SUS score depending on the type of condition used, $\chi^2(2)$ =12.441, p = 0.002. Median (IQR) for the SUS scores of voice only, mouse + voice, and gaze + voice conditions were 62.5 (45.0 to 75.0), 85.0 (67.5 to 87.5) and 82.5 (72.5 to 90.0), respectively.

Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at p < 0.017.

There were no significant differences between the gaze + voice and mouse + voice conditions (Z = -0.787, p = 0.431), despite an overall higher mean SUS score for the Gaze condition (median = 85 and 82.6 respectively) [see Figure 5.6]. However, there was a statistically significant lower mean SUS score for the voice only condition as compared to the mouse + voice (Z =-2.559, p = 0.010) or gaze + voice (Z = -2.502, p = 0.012) conditions.



[Figure 5.6] Box plot of System Usability Scale scores

5.3.3 MEASURE OF SOCIAL PRESENCE

The measure of social presence was carried out using a validated questionnaire [54] consisting of a set of 24 questions grouped into four sections:

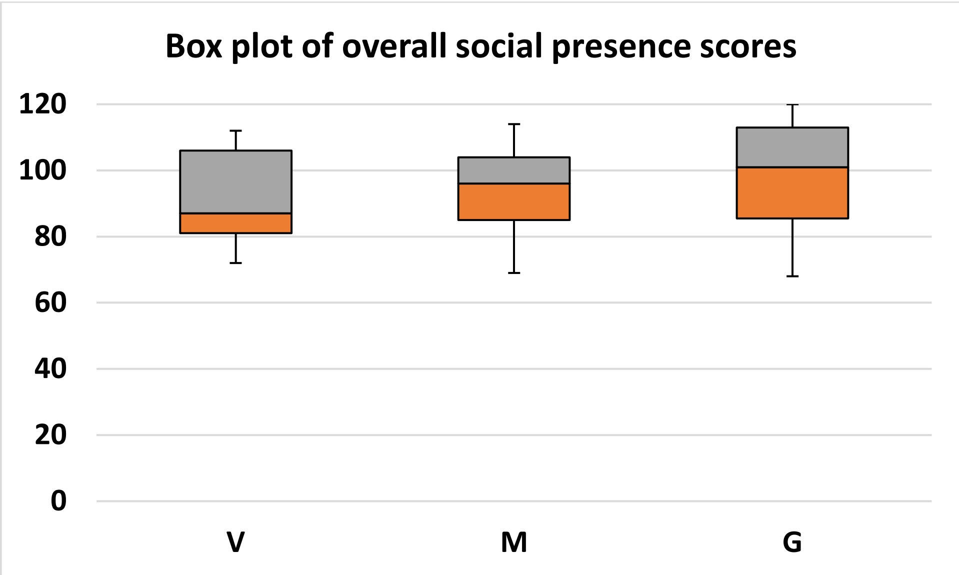
- Co-presence
- Attention allocation
- Perceived message understanding
- Perceived emotional understanding

The participants rated each of the 24 questions from 1 (Strongly disagree) to 5 (Strongly agree).

- For each of the negative questions, the scores were subtracted from 6 so to measure it from 1 to 5
- The scores for positive questions were left the same
- Maximum score for the feedback questionnaire was 120 (24 x 5)

OVERALL COLLABORATIVE EXPERIENCE

Comparing to the results from the task completion time and SUS, there was no significant difference even with the voice condition while using the interface. The box plot of the social presence scores [Figure 5.11] show that although there is a minor higher score in favour of the gaze + voice condition, the participants rated all three conditions almost even. A detailed discussion on why the participants might have rated them almost equally has been included in the following chapters.



[Figure 5.7] Box plot of overall social presence test results scores

The mean scores of the entire Collaborative Experience Survey (N=15) for V, M, and G were 3.83, 3.95, and 4.12 respectively. An analysis of the collaborative experience survey scores using Friedman test found no significant difference between voice only, mouse + voice, and gaze + voice condition as well ($\chi^2(2)$ =3.448, p = 0.178) [Table 5.6].

		Percentiles		
	N	25th	50th (Median)	75th
V	15	3.33	3.62	4.42
M	15	3.46	4.00	4.33
G	15	3.50	4.21	4.75

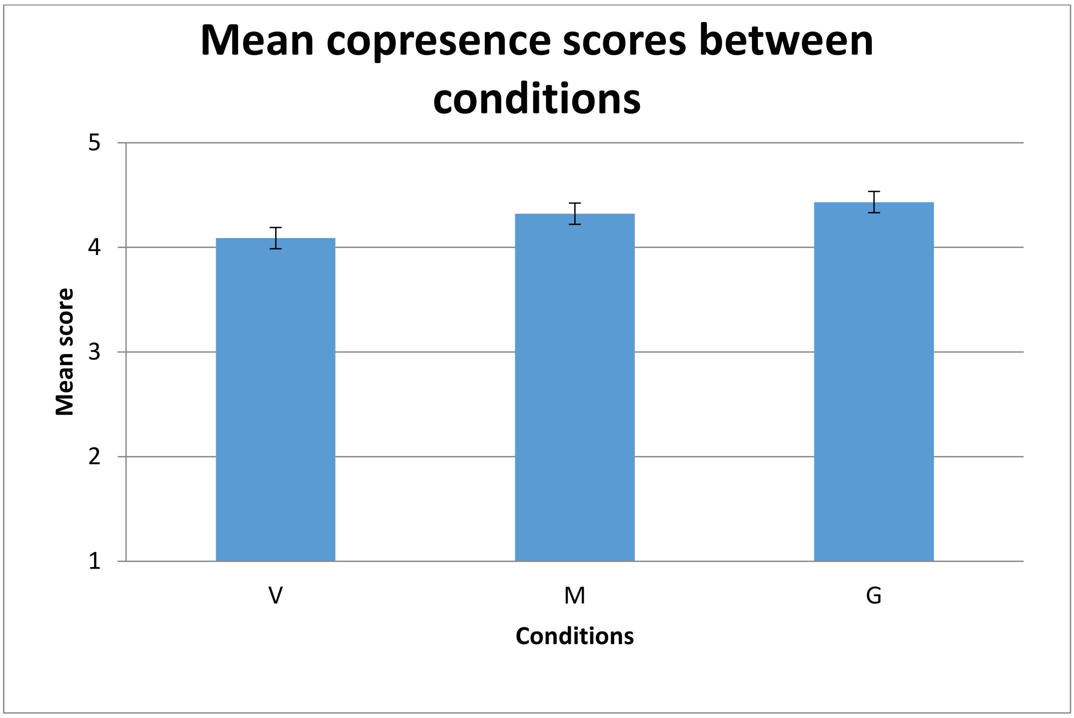
[Table 5.2] Descriptive Statistics of results from overall collaborative experience

CO-PRESENCE

The mean scores for copresence in the Collaborative Experience Survey (N=15) for V, M, and G were 4.09, 4.32, and 4.43 respectively [see Table 5.2]. Figure 5.6 shows a bar graph of the mean copresence scores between the three conditions.

	N	Percentiles		
		25th	50th (Median)	75th
V.Copresence	15	3.67	4.00	4.67
M.Copresence	15	3.83	4.50	5.00
G.Copresence	15	3.67	4.83	5.00

[Table 5.3] Descriptive statistics for results for co-presence



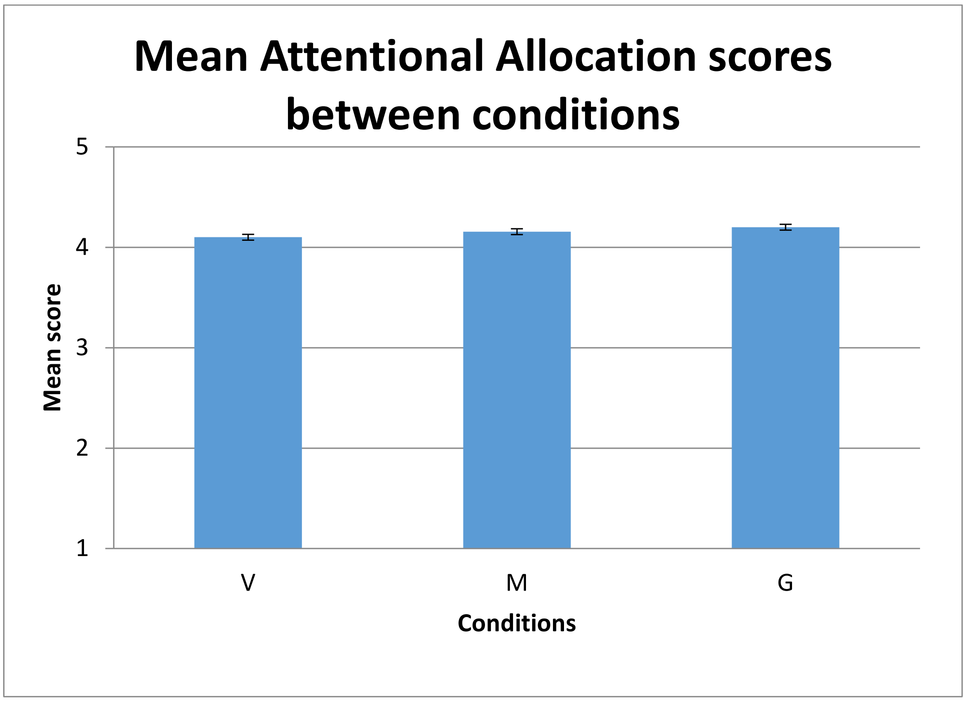
[Figure 5.8] Bar chart of mean co-presence scores between conditions with standard error (with standard error)

ATTENTION ALLOCATION

The mean scores for attentional allocation (N=15) for V, M, and G were 4.10, 4.16, and 4.20 respectively [see Table 5.3]. Figure 5.7 shows a bar graph of the mean attention allocation scores between the three conditions

	N	Percentiles		
		25th	50th (Median)	75th
V.AttentionalAllocation	15	3.67	4.00	4.83
M.AttentionalAllocation	15	3.67	4.167	4.667
G.AttentionalAllocation	15	3.34	4.34	5.00

[Table 5.4] Descriptive Statistics of results from attention allocation



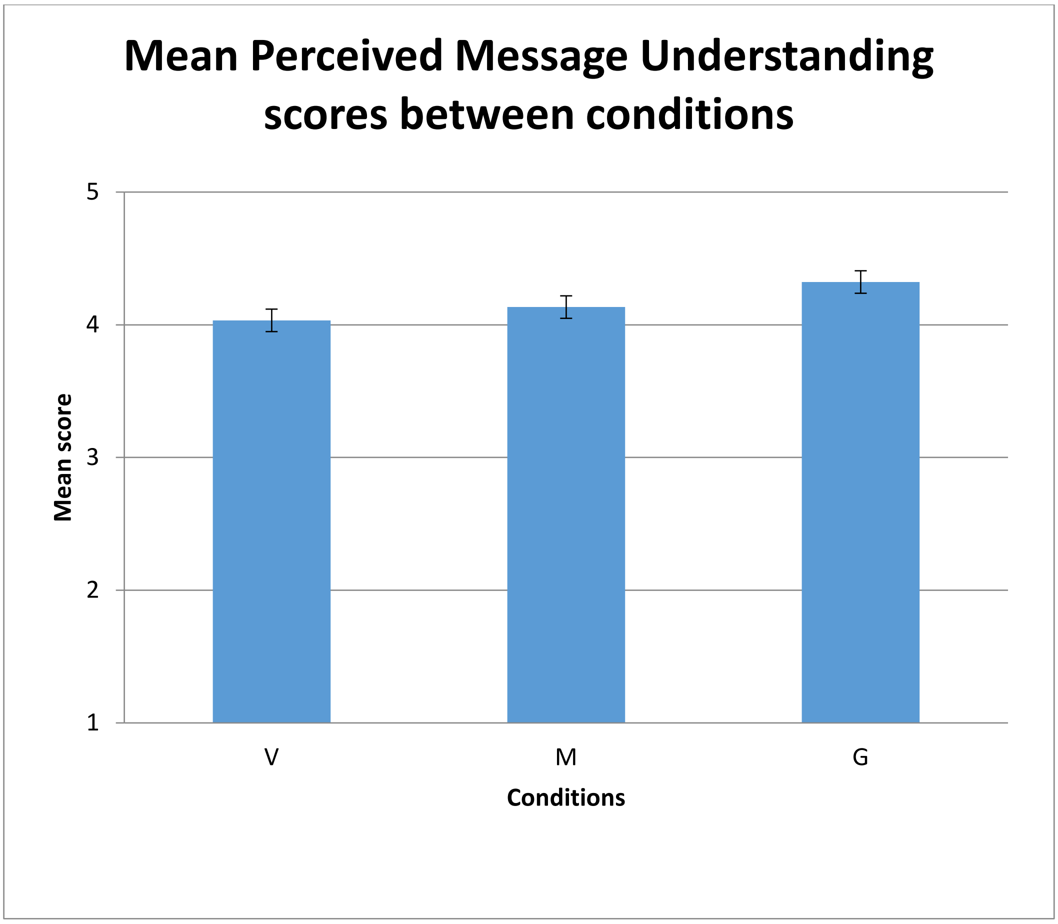
[Figure 5.9] Bar chart of mean attention allocation scores between conditions (with standard error)

PERCEIVED MESSAGE UNDERSTANDING

The mean scores for perceived message understanding (N=15) for V, M, and G were 4.03, 4.13, and 4.32 respectively [see Table 5.4]. Figure 5.8 shows the bar chart of the mean perceived message understanding scores between the three conditions.

Descriptive Statistics of results from perceived message understanding				
	N	Percentiles		
		25th	50th (Median)	75th
V.PerceivedMessageUnderstanding	15	3.33	4.00	5.00
M.PerceivedMessageUnderstanding	15	3.50	4.33	4.67
G.PerceivedMessageUnderstanding	15	3.67	4.33	5.00

[Table 5.5] Descriptive Statistics of results from perceived message understanding



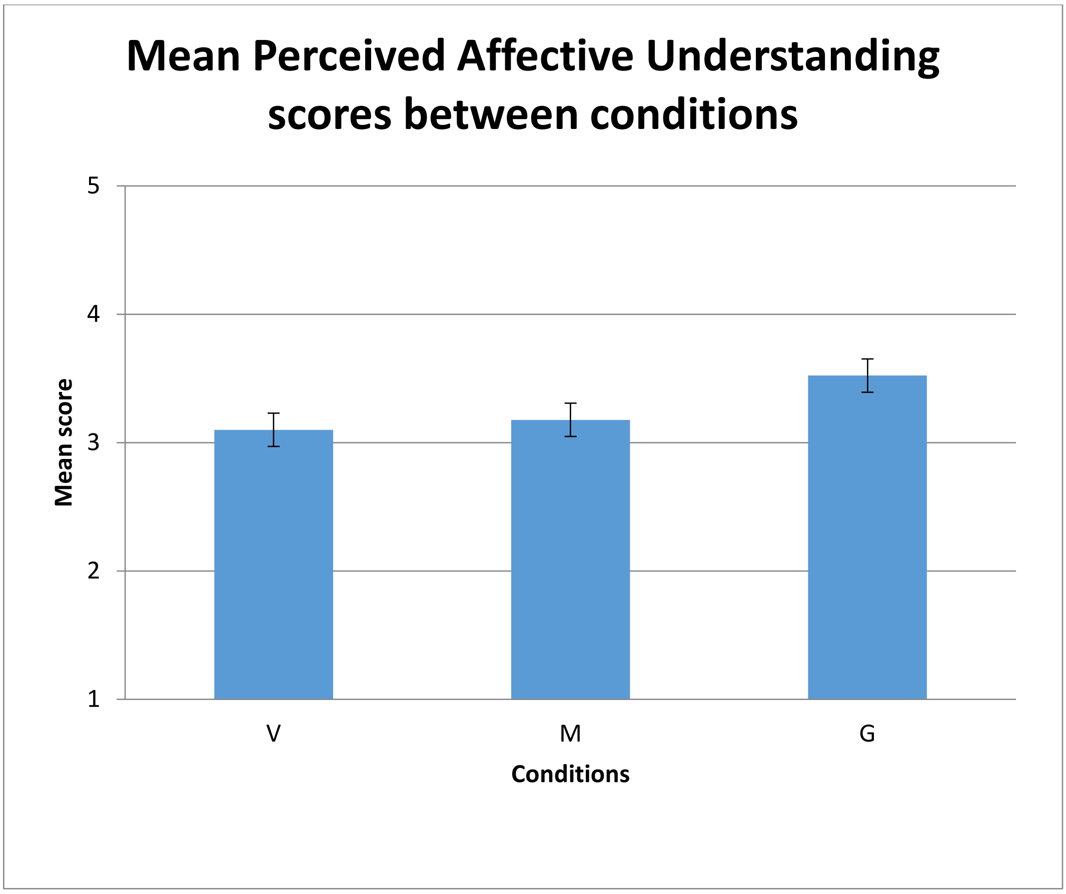
[Figure 5.10] Bar chart of mean perceived message understanding scores between conditions (with standard error)

PERCEIVED AFFECTIVE UNDERSTANDING

The mean scores for perceived affective understanding (N=15) for V, M, and G were 3.10, 3.18, and 3.52 respectively [see Table 5.5]. Figure 5.9 shows a bar graph of the mean perceived affective understanding scores between the three conditions.

		Percentiles		
	N	25th	50th (Median)	75th
V.PerceivedAffectiveUnderstanding	15	2.33	2.83	4.00
M.PerceivedAffectiveUnderstanding	15	2.33	3.00	4.00
G.PerceivedAffectiveUnderstanding	15	2.67	3.67	4.67

[Table 5.6] Descriptive Statistics of results from perceived affective understanding



[Figure 5.11] Bar chart of mean perceived affective understanding scores between conditions (with standard error)

An analysis of the data was conducted using Friedman Test found no significant difference found between voice only , mouse + voice, and gaze + voice conditions in Copresence ($\chi^2(2)$ =3.714, p = 0.156), Attentional Allocation ($\chi^2(2)$ =0.857, p = 0.651), Perceived Message Understanding ($\chi^2(2)$ =2.811, p = 0.245), Perceived Affective Understanding ($\chi^2(2)$ =3.957, p = 0.138).

5.3.4 RANKING THE CONDITIONS

At the end of the experiment, the participants were requested to rank the different conditions for the following questions:

Conditions: Voice only (V), Mouse + voice cues (M), Gaze + voice cues (G)

Rank the conditions (1,2,3) *(1 - best, 3 – worst)*	**V**	**M**	**G**
At helping you **complete** the task			
At making you feel **connected** with your partner			
At helping you stay **focused** on the task			
At making you feel that you were **present** with you partner			
For you (or the partner) to **know** that the partner (or you) **needed assistance**?			
At helping you **understand** the partner's message?			

[Table 5.7] Ranking questionnaire that was included in the post-experiment questionnaire

Q1. At helping you complete the task

An analysis of the data was conducted using Friedman Test found significant difference between voice only, mouse + voice, and gaze + voice conditions in ranking for Q1 ($\chi^2(2)$ =17.733, $p < 0.001$). Median (IQR) for the ranking of voice only, mouse + voice, and gaze + voice conditions were 3 (3 to 3), 2 (1 to 2) and 1 (1 to 2), respectively.

Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < 0.017$.

There was significant difference between the mouse + voice and voice only conditions (Z = -2.504, p = 0.012), voice and gaze + voice conditions (Z = -3.508, p < 0.001) and between gaze + voice and mouse + voice conditions (Z = -2.055, p = 0.040).

Q2. At making you feel connected with your partner

An analysis of the data was conducted using Friedman Test found significant difference between voice only, mouse + voice, and gaze + voice conditions in ranking for Q2 ($\chi^2(2) = 8.933$, p = 0.011). Median (IQR) for the ranking of voice, mouse + voice, and gaze + voice conditions were 3 (2 to 3), 2 (2 to 3) and 1 (1 to 2), respectively.

There was no significant difference between the mouse + voice and voice only conditions (Z = -0.894, *p* = 0.371). There was a significant difference between voice and gaze + voice conditions (Z = -2.464, *p* = 0.014), also there was significant difference between the gaze + voice and mouse + voice conditions (Z = -2.399, *p* = 0.016).

Q3. At helping you stay focused on the task

An analysis of the data was conducted using Friedman Test found significant difference between voice only, mouse + voice, and gaze + voice conditions in ranking for Q3 ($\chi^2(2) = 8.533$, p = 0.014). Median (IQR) for the ranking of voice only, mouse + voice, and gaze + voice conditions were 3 (2 to 3), 2 (2 to 2) and 1 (1 to 2), respectively.

There was no significant difference between the mouse + voice and voice conditions (Z = -1.602, *p* = 0.109). There was a significant difference between voice and gaze + voice conditions (Z = -2.464, *p* = 0.014), also there was no significant difference between the gaze + voice and mouse + voice conditions (Z = -1.602, *p* = 0.109).

Q4. At making you feel that you were present with you partner

An analysis of the data was conducted using Friedman Test found significant difference between voice only, mouse + voice, and gaze + voice conditions in ranking for Q4 ($\chi^2(2) = 10.133$, p = 0.006). Median (IQR) for the ranking of voice only, mouse + voice, and gaze + voice conditions were 3 (2 to 3), 2 (2 to 2) and 1 (1 to 2), respectively.

There was no significant difference between the mouse + voice and voice conditions (Z = -0.471, *p* = 0.637). There was a significant difference between voice and gaze + voice conditions (Z = -2.464, p = 0.014), also there was significant difference between the gaze + voice and mouse + voice conditions (Z = -2.450, p = 0.014).

Q5. For you (or the partner) to know that the partner (or you) needed assistance?

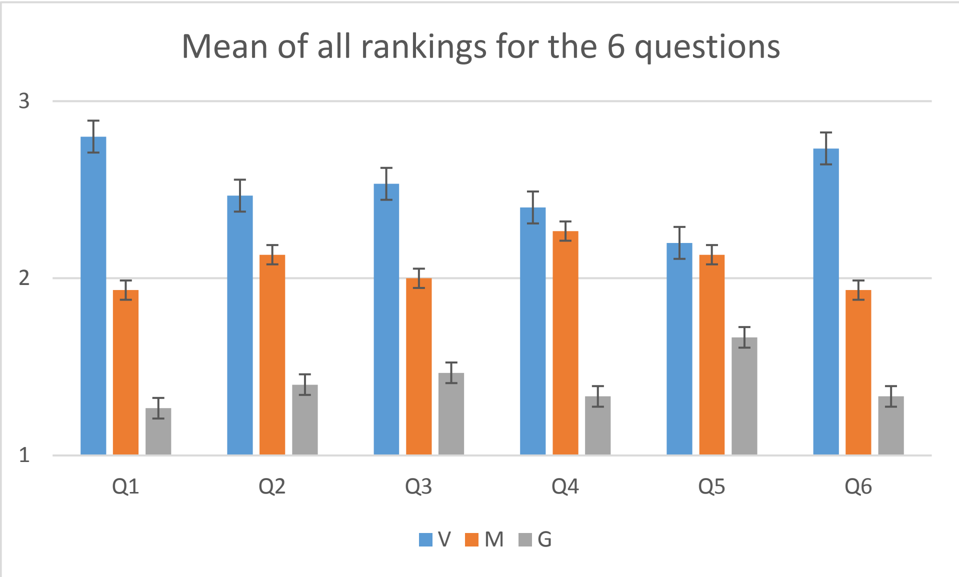
An analysis of the data was conducted using Friedman Test found no significant difference between voice only, mouse + voice, and gaze + voice conditions in ranking for Q5 ($\chi^2(2) = 2.533$, p = 0.282). Median (IQR) for the ranking of voice only, mouse + voice, and gaze + voice conditions were 3 (3 to 3), 2 (1 to 2) and 1 (1 to 2), respectively.

Since there was no significant difference found in Friedman Test, no post-hoc tests were made.

Q6. At helping you understand the partner's message?

An analysis of the data was conducted using Friedman Test found significant difference between voice, mouse + voice, and gaze + voice conditions in ranking for Q6 ($\chi^2(2)$ = 14.800, p = 0.001). Median (IQR) for the ranking of voice only, mouse + voice, and gaze + voice conditions were 3 (3 to 3), 2 (1 to 2) and 1 (1 to 2), respectively.

There was significant difference between the mouse + voice and voice conditions (Z = -2.387, p = 0.016), also there was significant difference between voice and gaze + voice conditions (Z = -3.188, p = 0.001). However, there was no significant difference between the gaze + voice and mouse + voice conditions (Z = -1.937, p = 0.053).



*The rankings have been displayed from 1(highest) to 3(lowest).

[Figure 5.12] Bar chart of overall ranking results from the remote user's feedback (with error bars)

Overall, there was significant results in favour of the gaze + voice condition over voice and mouse + voice conditions, in completing the task (Q1), in helping the participants feel connected during the collaboration process (Q2) and making them feel present with their partners (Q4). Also, there was significant difference within the three conditions at helping them stay focused in the task (Q3) and at understanding the partner's message (Q6). However, the results found no significant difference between the gaze + voice and mouse + voice conditions in Q3 and Q6. There seemed to be no significant difference in them knowing if their partner needed any more assistance (Q5). However, there was a consistent trend showing that the participants preferred the gaze + voice and mouse + voice conditions over the voice condition in their collaborative task. A detailed analysis of the ranking results have been discussed in the following chapters.

5.3.5 QUALITATIVE FEEDBACK

After completing the instructions for all three conditions the participants were asked to fill out a post-experiment questionnaire. The questionnaire consisted of a set of statements which the participants were requested to rank the different conditions. They were also requested to provide some feedback on their most preferred condition and their least favourite condition in the post-experiment questionnaire and give their comments on what could be improved in the interface. The users were also asked the reason for their rating behind each condition. The top answers from the user feedback with respect to their favourite conditions have been summarised below.

WHAT DID YOU LIKE THE MOST IN THE CONDITION YOU RANKED BEST?

For Mouse + voice cues:

- Mouse was clear and simple.
- It was possible to communicate the instructions easily.
- Being able to see the environment and physically guide the actions was.

To quote one of the answers from the participant who ranked the mouse condition the highest, *"M tops well into familiar behaviours, gave all of the possible information easily".*

For Gaze + voice cues:

- The instructions were spontaneous using gaze cues and easy to convey the instructions to the local users.
- The advantage of having their partners visually look at where the remote users were looking made the gaze + voice condition a preferred choice.
- The fact that there was less vocal communication made the collaboration process using the gaze + voice cues easy to use.

Quoting a few of the participants who preferred the gaze cues; *"The ease with which my partner was able to pick up on my cues based on where I was looking", "G was easy to communicate where I was looking once I had located it", "G didn't need to give as much detail because I could just look at it".*

For voice cues: The voice condition was not ranked best by any of the participants.

WHAT DID YOU DISLIKE THE MOST IN THE CONDITION YOU RANKED LEAST?

For voice cues:

- Just using voice was hard to provide a context to the local users and made it hard to explain the exact location of the action button.
- For non-native English speakers, it was difficult to communicate the instructions considering they had to make it simple for their partners to understand.
- It took an extra effort to instruct their partners when the remote users were not sure of a description for the action button.

To quote a few of the feedback of why the participants thought the voice cues were the least favourite condition; *“I had to describe using only voice in simple language as my partner is a native foreign speaker language”, “It could get confusing using only voice cause sometimes you don’t know what the thing on the front is called”, “It was hard to communicate where exactly I wanted to focus using my words.”*

For mouse + voice cues:

- Since the local users were constantly moving to identify the action button in the three screen projected control panel, it was difficult for the remote users to move their mouse along with their partner’s movement.

To quote a few of the participants feedback; *“M - The system seemed difficult to use because when my partner moved it was in the wrong place”, “The extra step and effort of dragging the mouse.”*

For gaze + voice cues:

- The local users performed actions as soon as the remote users stalled their gaze at a particular point even before the instructions were made.

- Gaze movements were not accurate in a few places where there were action buttons close to each other. This made it distracting for their partners.

To quote a few of the participants feedback; "*E - had to focus too much on the mechanics rather than the content of the task. Having to keep eyes steady prevented scanning", "In 'G' moving the eyes are distracting the partner a little bit."*

PLEASE STATE WHAT COULD BE IMPROVED IN THE 3 CONDITIONS

- Stability of the eye tracking since the gaze movements were jittery
- Locking the position of the pointers in the mouse + voice and gaze + voice conditions - If the participants have identified the desired action button, it would be easier if they can lock the position of the pointer. This way even if their partners are moving, the pointer would still be locked in the identified position.

A few of the participant's feedback on the improvement of the prototype have been listed below:

- *"A simple addition to 'M' would be to have a toggleable marker (i.e. on click hold/toggle with right key) to prevent distraction on larger transitions. A more complex addition would be for the dot to turn into a rotating/dynamic arrow on longer mouse drags."*
- *"In the mouse system you could use a visual representation of your clicks. In the glass you could have a feature that locked the highlighted area so that you could point that place while you look somewhere else."*
- *"The quality of camera was bad. It was difficult to figure out colours and symbols."*

ANY FURTHER COMMENTS

A selection of the user feedback on how they liked the prototype and further comments on the improvement on the system have been listed below:

- *“It was awesome”*
- *“Maybe only pseudo or randomise conditions i.e have 2 actions on each of the left/right screens and one on the centre, then randomise presets across condition and internally.”*
- *“Very useful in construction sector”*
- *“I would recommend the gaze tracker for futuristic performances in various industry”*
- *“I need to slump downward for the camera to recognize me. If the pink detected onto an item initially that would have been easier to use and yes/no to say it had locked on the right one.”*

5.4 Conclusion

The results from the comparison of the task completion time showed that the participants were able to complete the task faster using gaze + voice cues over mouse + voice and only voice. This proves Hypothesis 1 that the gaze + voice cues are most effective in completing a given task in a collaborative environment over video conferencing.

The feedback from the users on SUS and the overall ranking show a mostly favourable results towards the usage of gaze + voice cues during the collaborative process. The use of just voice cues for the interaction was significantly lower when compared to the other two conditions. This was also evident from the qualitative feedback from the users in the post experiment questionnaire. From the recorded time for performing the various conditions, the participants took a comparatively lower time for the task completion with the use of mouse and gaze + voice cues in comparison to voice cues. However, since there was no significant difference between the mouse + voice condition and the gaze + voice condition in their social presence survey and the SUS survey, Hypothesis 2 (H2) has to be rejected.

From the comments on the qualitative feedback provided by the users, it is clear that the users enjoyed the collaborative process. They felt that it would be a valuable addition to have mouse or gaze cues added to their regular communication process when performing a task together. Some of the participants from the construction businesses felt strongly that the added conditions would be a great asset in their field while performing tasks such as surveying, reporting and construction management, to name a few.

Since there was no significant results with the three conditions on the social presence level, Hypothesis 3 has to be rejected. The users felt connected with their partners with all three conditions since they were interacting with each other with not much distractions in the collaborative process.

The results from the rankings for the three conditions showed a consistent trend in favour of the gaze + voice condition over the mouse + voice and voice conditions. Thus, Hypothesis 4 has been proved true. The participants felt that the gaze + voice cues helped them feel present and connected with their partners in the collaborative process.

Adding on to the feedback from the remote users, the local users were also requested to provide their comments in the post experiment questionnaire. Similar to the results from the remote users, the results local users also had an overall trend in favour of the gaze + voice condition over the mouse + voice and voice conditions. The local users felt that they were able to receive the message well from their partners using the gaze + voice and mouse + voice cues. They also felt that it was time consuming when their partners tried to instruct them vocally while identifying the action buttons. To quote one of the participants who ranked in favour of the gaze + voice condition, he said, "*It was easy to perform the action with the help of the gaze pointer. Time saving*." He also said, "*It took a lot of time to search the buttons that my partner was trying to explain how the button looks.*"

The next section elaborates more on the results followed by a few pointers on future work.

6

Discussion

The overall results with the task completion time, qualitative feedback and ranking supports Hypothesis 1 and 4, and related research suggests that the gaze + voice condition is significantly preferred over mouse + voice and voice conditions. This tells us that gaze + voice condition has potential and would be a favourable option to replace the mouse and voice conditions for future remote collaboration.

The results from the SUS survey do not support the hypothesis 2 that gaze + voice condition is better than the mouse + voice condition. The results from the social presence questionnaire rejected the hypothesis 3. However, related research suggests that the gaze + voice and mouse + voice conditions are significantly better than the voice cues.

6.1 Reasons for differing results from earlier research

This study tested gaze tracking on remote users, where local users could see where the remote user was looking at, but Gupta's study [8] tested gaze tracking on local users, where remote users could see what the local user was looking at.

Perhaps remote users have similar experiences for V/M/G because they are not the ones who have to complete the task, while local users have difference in experience because they are the ones who have to complete the task. (i.e. Local users feel better when they

finish faster/easier, but remote users don’t care whether the local users finish faster/slower or find it easier to complete the task or not).
Gupta’s study [8] used a task that required cognitive thinking while this study used a task that did not require cognitive thinking, but required searching (matching) for the right button shape/pattern. Perhaps the lack of cognitive thinking resulted in remote users feeling similar for the collaborative experiences between V/M/G (i.e. boring task result in similar experience rating). Future study should replicate this study with Gupta’s study task to eliminate this as a possible variable.

Another major difference in Gupta’s study is that it did not require local user to look around (turn his head left/right), but this study did. Perhaps the constant movement of the environment screen resulted in remote users feeling similar for the collaborative experiences between V/M/G. (i.e. nauseating movements caused headaches and therefore there was similar experience rating). Future study should replicate this study with Gupta’s study’s task to eliminate this as a possible variable.

Also, Gupta’s study had the local user in a confined space while this study had local users in a free space and have to move around to touch buttons. For the purpose of testing local users, future study should replicate this study with Gupta’s study environment to eliminate this as a possible variable.

6.2 Reasons for having similar social presence survey results

The initial study design wanted to physically isolate the remote and local users, but unfortunately the technology did not allow it (the glasses did not work at a distance from the local user space). Remote users were therefore seated behind the remote users, with the front of their view blocked to ensure they are unable to see the screen in real life. However, their surrounding walls were not blocked out. This might not be sufficient to create the illusion that the local and remote users were not in the same space.

One of the most essential benefits of gaze condition in the future would be that the remote users would be able to move around and without having to sit in front of the laptop screen. But for this study, the remote users had to sit in front of the laptop screen as the eye tracker could not be moved. Thus this benefit was not demonstrated in this study.

As the survey was self-declared, it might not be as accurate as tracking their physiological changes. Future research should use emotion detection software or track their galvanic skin responses to obtain more accurate results.

The results do not show significant difference between the different conditions in the tests analysed. There is, however, a difference observed, but it is not statistically significant. Results for task completion times, SUS survey, rankings all show trends in favour of gaze + voice as compared to voice or mouse + voice. Perhaps this survey requires more participants to obtain significant results. For this study, time was a limitation.

Many participants already knew each other beforehand, and this might cause biased results as they feel already connected to each other. Most participants had the same ethnicity/were from the same country, and this might cause biased results as they feel already connected to each other. The task not indicating any success or failure also might have caused subtle difference.

7

Conclusion and Future work

My research investigated some of characteristics of the effects of visual cues during a remote collaborative process. Nevertheless, there is a lot of scope on how these could be extended and reiterated for future work. This chapter elaborates on some of the potential outcomes for future research and how technology in future could be beneficial pertaining to this sector of research.

One of the most essential benefits of gaze cues in the future would be that the remote users would be able to move around and without having to sit in front of the laptop screen. But for this study, the remote users had to sit in front of the laptop screen as the eye tracker could not be moved. Thus this benefit was not demonstrated in this study.

As the survey was self-declared, it might not be as accurate as tracking their physiological changes. Future research should use emotion detection software or track their galvanic skin responses to obtain more accurate results.

Even though there was no significant difference in the SUS and social presence score results between the gaze + voice and mouse + voice conditions, the overall results showed a strong trend in favour of the gaze + voice condition. The task completion time results also showed that the participants were able to perform the tasks faster using the gaze + voice cues. Also, the results from the rankings for each condition showed a consistent trend in favour of the gaze + voice cues over the mouse + voice and voice conditions. There is a difference in the results but the sample size might be too small to detect a significant difference. As this study was limited by time and budget, future studies could test a bigger participant pool.

The results from the individual ranking for the three conditions showed that the gaze + voice condition was preferred over voice in making them feel connected, completing their task, understanding their partner's message and at making the participants feel present in the collaboration process. The results also showed a consistent trend in favour of gaze + voice over mouse + voice and only voice. On analysis of their qualitative feedback, the reason for their overall preference for gaze + voice cues was the ease-of-use and the concept of visually sharing their gaze location in a real-time environment.

My research did not investigate all the possible interfaces that could be used during the collaborative process. Although the Vuzix 1200XLD proved to be one of the better choices for HMD at the local users end, there could be more options for a more convenient and hassle free display. Much of the qualitative feedback from the local users that they were a little uncomfortable using a device with a lot of connections attached to the device. With the advancements of technologies such as Oculus and Google Glass, we could potentially avoid using multiple wired connections for the HMD. Transmitting information (camera feed) through a wireless HMD could potentially reduce the burden of the users device plugged into multiple connections and USB extensions. It would be exciting to equate other HMDs over the Vuzix system and compare results using different options.

User studies on similar research could extend from my study with interactive prototypes, using eye trackers for both the local users and the remote users in a video conferencing system and using different conditions to mimic a real-life scenario such as performing a plumbing task, simulate an army mission, reporting a construction field work, etc.

In the process of forming this thesis, my understanding has developed about user expectations and essentials in sharing information remotely. From the overall results of the experiment, it was learnt that users want to feel connected in a collaborative process and using gaze + voice cues has a positive impact on this. Future development of gaze cues within the user interface could benefit many industries including the construction sector, IT technical support and trade services. Gaze cues hold promise for many sectors where the development of a shared and common understanding is essential.

R

References

[1] J. V. Wertsch and P. Tulviste, “L. S. Vygotsky and contemporary developmental psychology,” *Dev. Psychol.*, vol. 28, no. 4, pp. 548–557, 1992.

[2] A. A. Sawchuk, E. Chew, R. Zimmermann, C. Papadopoulos, and C. Kyriakakis, “From Remote Media Immersion to Distributed Immersive Performance,” in *Proceedings of the 2003 ACM SIGMM Workshop on Experiential Telepresence*, New York, NY, USA, 2003, pp. 110–120.

[3] B. Schneider and R. Pea, “Real-time mutual gaze perception enhances collaborative learning and collaboration quality,” *Int. J. Comput.-Support. Collab. Learn.*, vol. 8, no. 4, pp. 375–397, Oct. 2013.

[4] M. B. Neider, X. Chen, C. A. Dickinson, S. E. Brennan, and G. J. Zelinsky, “Coordinating spatial referencing using shared gaze,” *Psychon. Bull. Rev.*, vol. 17, no. 5, pp. 718–724, Oct. 2010.

[5] X. C. Mark B Neider, “Coordinating spatial referencing using shared gaze,” *Psychon. Bull. Amp Rev.*, vol. 17, no. 5, pp. 718–24, 2010.

[6] C. Moore and P. Dunham, *Joint Attention: Its Origins and Role in Development*. Psychology Press, 2014.

[7] S. E. Brennan, X. Chen, C. A. Dickinson, M. B. Neider, and G. J. Zelinsky, “Coordinating cognition: The costs and benefits of shared gaze during collaborative search,” *Cognition*, vol. 106, no. 3, pp. 1465–1477, Mar. 2008.

[8] K. Gupta, “Wearable Tools for Affective Remote Collaboration,” 2015.

[9] “Skype | Free calls to friends and family.” [Online]. Available: https://www.skype.com/en/. [Accessed: 13-Jul-2016].

[10] T. Inc, “Periscope,” *Periscope*. [Online]. Available: https://www.periscope.tv. [Accessed: 13-Jul-2016].

[11] “FaceTime - Make video calls from your Mac - Apple (NZ).” [Online]. Available: http://www.apple.com/nz/mac/facetime/. [Accessed: 13-Jul-2016].

[12] “Live | Facebook.” .

[13] D. Derks, A. H. Fischer, and A. E. R. Bos, “The role of emotion in computer-mediated communication: A review,” *Comput. Hum. Behav.*, vol. 24, no. 3, pp. 766–785, May 2008.

[14] R. W. Picard, *Affective Computing*. 1995.

[15] R. Müller, J. R. Helmert, S. Pannasch, and B. M. Velichkovsky, “Gaze transfer in remote cooperation: Is it always helpful to see what your partner is attending to?,” *Q. J. Exp. Psychol.*, vol. 66, no. 7, pp. 1302–1316, Jul. 2013.

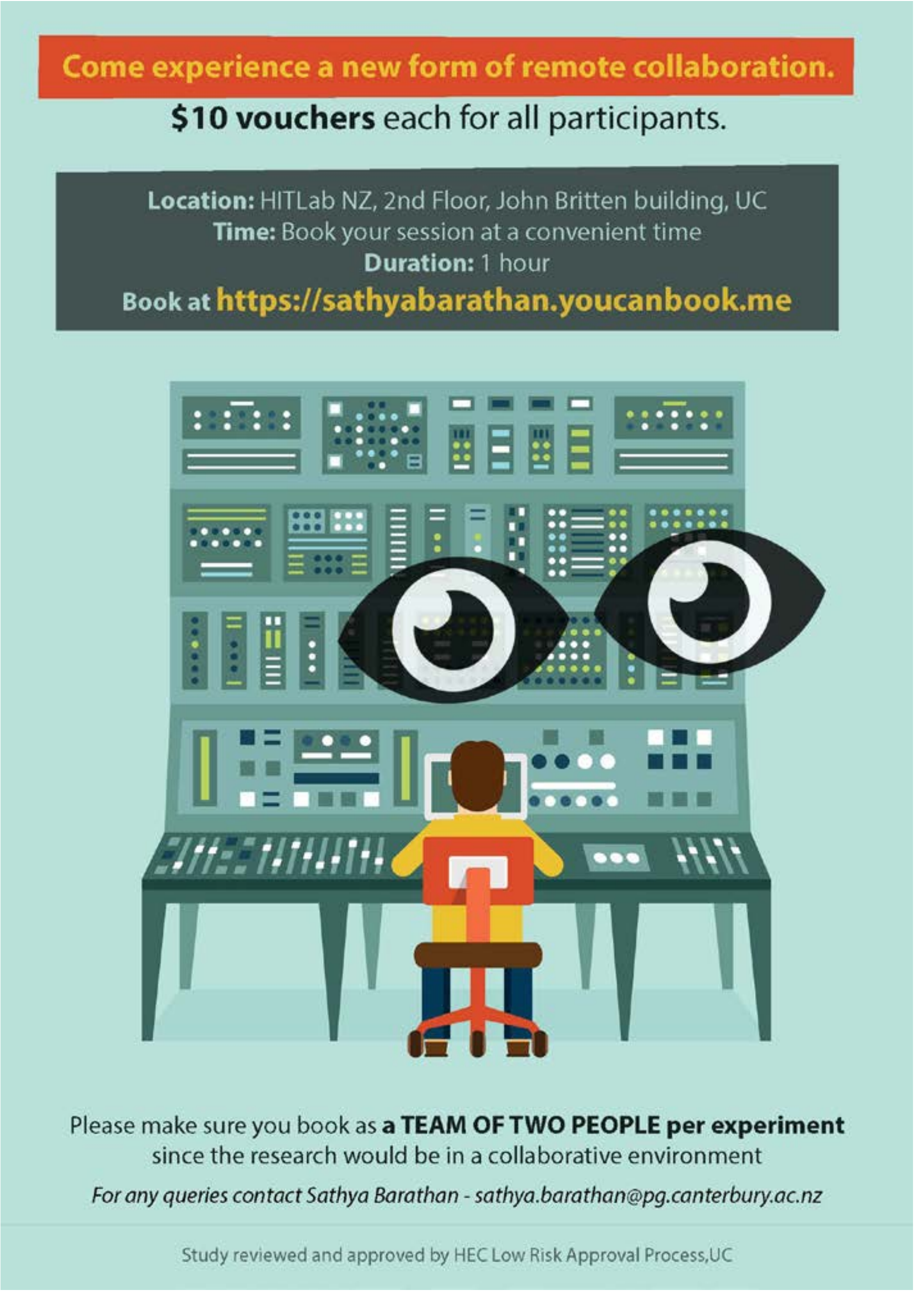
[16] J. Gemmell, K. Toyama, C. L. Zitnick, T. Kang, and S. Seitz, “Gaze Awareness for Video-Conferencing: A Software Approach,” *IEEE Multimedia*, vol. 7, no. 4, pp. 26–35, 2000.

[17] R. E. Kraut, M. D. Miller, and J. Siegel, “Collaboration in Performance of Physical Tasks: Effects on Outcomes and Communication,” in *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work*, New York, NY, USA, 1996, pp. 57–66.
[18] J. Siegel, R. E. Kraut, B. E. John, and K. M. Carley, “An Empirical Study of Collaborative Wearable Computer Systems,” in *Conference Companion on Human Factors in Computing Systems*, New York, NY, USA, 1995, pp. 312–313.
[19] M. Billinghurst, J. Bowskill, N. Dyer, and J. Morphett, “An evaluation of wearable information spaces,” in *Virtual Reality Annual International Symposium, 1998. Proceedings., IEEE 1998*, 1998, pp. 20–27.
[20] M. Billinghurst, S. Weghorst, and T. Furness, “Shared space: An augmented reality approach for computer supported collaborative work,” *Virtual Real.*, vol. 3, no. 1, pp. 25–36.
[21] P. Kauff and O. Schreer, “An Immersive 3D Video-conferencing System Using Shared Virtual Team User Environments,” in *Proceedings of the 4th International Conference on Collaborative Virtual Environments*, New York, NY, USA, 2002, pp. 105–112.
[22] R. Vertegaal, I. Weevers, C. Sohn, and C. Cheung, “GAZE-2: Conveying Eye Contact in Group Video Conferencing Using Eye-controlled Camera Direction,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2003, pp. 521–528.
[23] A. Tang, C. Neustaedter, and S. Greenberg, “VideoArms: Embodiments for Mixed Presence Groupware,” in *People and Computers XX — Engage*, N. B.-K. Bs. AKC MSc,, PGCAP, A. B. B. CITP MA,, MBCS, CEng, P. C. B. PGCertHE MA, and L. N. Bs. MSc, Eds. Springer London, 2007, pp. 85–102.
[24] P. Schwartz, L. Bricker, B. Campbell, T. Furness, K. Inkpen, L. Matheson, N. Nakamura, L.-S. Shen, S. Tanney, and S. Yen, “Virtual Playground: Architectures for a Shared Virtual World,” in *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, New York, NY, USA, 1998, pp. 43–50.
[25] K. Inkpen, B. Taylor, S. Junuzovic, J. Tang, and G. Venolia, “Experiences2Go: Sharing Kids’ Activities Outside the Home with Remote Family Members,” in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, New York, NY, USA, 2013, pp. 1329–1340.
[26] S. d. Scott, R. l. Mandryk, and K. m. Inkpen, “Understanding children’s collaborative interactions in shared environments,” *J. Comput. Assist. Learn.*, vol. 19, no. 2, pp. 220–228, Jun. 2003.
[27] C. Egido, “Video Conferencing As a Technology to Support Group Work: A Review of Its Failures,” in *Proceedings of the 1988 ACM Conference on Computer-supported Cooperative Work*, New York, NY, USA, 1988, pp. 13–24.
[28] M. Bauer, T. Heiber, G. Kortuem, and Z. Segall, “A collaborative wearable system with remote sensing,” in *Second International Symposium on Wearable Computers, 1998. Digest of Papers*, 1998, pp. 10–17.
[29] R. E. Kraut, S. R. Fussell, and J. Siegel, “Visual Information As a Conversational Resource in Collaborative Physical Tasks,” *Hum-Comput Interact*, vol. 18, no. 1, pp. 13–49, Jun. 2003.
[30] S. R. Fussell, R. E. Kraut, and J. Siegel, “Coordination of Communication: Effects of Shared Visual Context on Collaborative Work,” in *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, New York, NY, USA, 2000, pp. 21–30.
[31] D. G. Armstrong, T. M. Rankin, N. A. Giovinco, J. L. Mills, and Y. Matsuoka, “A Heads-Up Display for Diabetic Limb Salvage Surgery A View Through the Google Looking Glass,” *J. Diabetes Sci. Technol.*, p. 1932296814535561, May 2014.
[32] B. A. Wells, “Head mounted display for miniature video display system,” US5003300 A, 26-Mar-1991.
[33] “MREAL Displays | MREAL Head Mounted Display - HM-A1 | Canon USA.” [Online]. Available: https://www.usa.canon.com/internet/portal/us/home/products/details/mixed-reality/mreal-displays/head-mounted-display. [Accessed: 13-Jul-2016].

[34] "Oculus Rift | Oculus." [Online]. Available: https://www3.oculus.com/en-us/rift/. [Accessed: 13-Jul-2016].
[35] "olympus head mounted display - Google Search." [Online]. Available: https://www.google.co.nz/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=olympus%20head%20mounted%20display. [Accessed: 13-Jul-2016].
[36] "Vuzix | View The Future." [Online]. Available: https://www.vuzix.com/Products/LegacyProduct/6. [Accessed: 11-Jul-2016].
[37] A. Duchowski, *Eye Tracking Methodology: Theory and Practice*. Springer Science & Business Media, 2007.
[38] J. Carletta, R. L. Hill, C. Nicol, T. Taylor, J. P. de Ruiter, and E. G. Bard, "Eyetracking for two-person tasks with manipulation of a virtual world," *Behav. Res. Methods*, vol. 42, no. 1, pp. 254–265, Feb. 2010.
[39] E. M. Reingold, N. Charness, M. Pomplun, and D. M. Stampe, "Visual Span in Expert Chess Players: Evidence From Eye Movements," *Psychol. Sci.*, vol. 12, no. 1, pp. 48–55, Jan. 2001.
[40] D. C. Richardson, R. Dale, and N. Z. Kirkham, "The Art of Conversation Is Coordination Common Ground and the Coupling of Eye Movements During Dialogue," *Psychol. Sci.*, vol. 18, no. 5, pp. 407–413, May 2007.
[41] J. Tao and T. Tan, "Affective Computing: A Review," in *Affective Computing and Intelligent Interaction*, J. Tao, T. Tan, and R. W. Picard, Eds. Springer Berlin Heidelberg, 2005, pp. 981–995.
[42] M. Strauss, C. Reynolds, S. Hughes, K. Park, G. McDarby, and R. W. Picard, "The HandWave Bluetooth Skin Conductance Sensor," in *Affective Computing and Intelligent Interaction*, J. Tao, T. Tan, and R. W. Picard, Eds. Springer Berlin Heidelberg, 2005, pp. 699–706.
[43] J. Healey and R. W. Picard, "StartleCam: a cybernetic wearable camera," in *Second International Symposium on Wearable Computers, 1998. Digest of Papers*, 1998, pp. 42–49.
[44] J. H. Janssen, E. L. van den Broek, and J. H. D. M. Westerink, "Personalized affective music player," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, 2009, pp. 1–6.
[45] M. Cherubini, R. de Oliveira, N. Oliver, and C. Ferran, "Gaze and Gestures in Telepresence: multimodality, embodiment, and roles of collaboration," *ArXiv10013150 Cs*, Jan. 2010.
[46] J. S. Lobmaier, B. P. Tiddeman, and D. I. Perrett, "Emotional expression modulates perceived gaze direction," *Emotion*, vol. 8, no. 4, pp. 573–577, 2008.
[47] G. Tarantino, C. Fazio, and R. M. Sperandeo-Mineo, "A pedagogical flight simulator for longitudinal airplane flight," *Comput. Appl. Eng. Educ.*, vol. 18, no. 1, pp. 144–156, Mar. 2010.
[48] "Logitech C920 HD Pro Webcam for Windows, Mac, and Chrome OS." [Online]. Available: http://www.logitech.com/en-nz/product/hd-pro-webcam-c920. [Accessed: 14-Jul-2016].
[49] "Basics | eyetribe-docs." [Online]. Available: http://dev.theeyetribe.com/general/. [Accessed: 04-Apr-2016].
[50] "Satellite P70-B | Satellite C | Satellite L | Satellite P | Satellite Z | Toshiba." [Online]. Available: http://www.toshiba.eu/laptops/product-filter/?sFamily=satellite-p,satellite-s&sSeries=satellite-p70-b. [Accessed: 11-Jul-2016].
[51] "Processing.org." [Online]. Available: https://processing.org/. [Accessed: 14-Jul-2016].
[52] A. S. for P. Affairs, "System Usability Scale (SUS)," 06-Sep-2013. [Online]. Available: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html. [Accessed: 14-Jul-2016].
[53] P. C. Harms and P. F. Biocca, "Internal Consistency and Reliability of the Networked Minds Measure of Social Presence," 2004.
[54] S. Kim, G. Lee, N. Sakata, and M. Billinghurst, "Improving co-presence with augmented visual communication cues for sharing experience through video conference," in *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2014, pp. 83–92.

Appendix

A



Come experience a new form of remote collaboration.
$10 vouchers each for all participants.
Location: HITLab NZ, 2nd Floor, John Britten building, UC
Time: Book your session at a convenient time
Duration: 1 hour
Book at https://sathyabarathan.youcanbook.me
Please make sure you book as a TEAM OF TWO PEOPLE per experiment
since the research would be in a collaborative environment
For any queries contact Sathya Barathan - sathya.barathan@pg.canterbury.ac.nz
Study reviewed and approved by HEC Low Risk Approval Process,UC

B

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

HUMAN ETHICS COMMITTEE

Secretary, Rebecca Robinson
Telephone: +64 03 364 2987, Extn 45588
Email: human-ethics@canterbury.ac.nz

Ref: HEC 2016/36/LR

2 June 2016

Sathya Kumar Barathan
HIT Lab NZ
UNIVERSITY OF CANTERBURY

Dear Sathya Kumar

Thank you for submitting your low risk application to the Human Ethics Committee for the research proposal titled "Generating an Efficient Remote Collaboration Environment Using Shared Gaze".

I am pleased to advise that the application has been reviewed and approved.

With best wishes for your project.

Yours sincerely

pp. R. Robinson

Jane Maidment
Chair, Human Ethics Committee

C

Human Interface Technology Laboratory New Zealand
Telephone: +64 3 364 2349
Email: sathya.barathan@pg.canterbury.ac.nz
May 25, 2016

Generating an efficient remote collaboration environment using shared gaze
Information Sheet for participants

My name is Sathya and I am a Masters student in the HITLab NZ, UC. The purpose of this research is to understand how the introduction of the instructor's perspective into a remote collaborative system improve the usability and experience of the remote sharing process. In my thesis, I would like to explore an interface wherein the video conferencing process could display the instructor's eye tracking. Hereby, making it simple for the person wearing the head-mounted display to clearly identify where the instructor is viewing in the video sharing process.

If you choose to take part in this study, your involvement in this project will be to communicate with your partner and perform a collaborative task together using shared gaze. You would be performing a similar task on three different conditions.

As a follow-up to this investigation, you will be answering a simple questionnaire about your experience using the interface at the end of each condition. At the end of the experiment you would complete a survey to get some feedback on the overall process.

The study may raise physical stress of standing and moving for performing experimental tasks, yet the level of stress will be within the scope of everyday life activity.

Participation is voluntary and you have the right to withdraw at any stage without penalty. You may ask for your raw data to be returned to you or destroyed at any point. If you withdraw, I will remove information relating to you. However, once analysis of raw data starts on June 15, 2016, it will become increasingly difficult to remove the influence of your data on the results.

The results of the project will be published as a thesis and in other venues with the data in an anonymised statistical form, but you will be assured of the complete confidentiality of data gathered in this investigation: your identity will not be made public without your prior consent. To ensure anonymity and confidentiality, the data would be available only to me. All data would be stored in a private space and be destroyed after five years. A thesis is a public document and will be available through the UC Library. Please indicate to the researcher on the consent form if you would like to receive a copy of the summary of results of the project.

The project is being carried out as a requirement for the completion of Masters in Human Interface Technology by Sathya Kumar Barathan under the supervision of Dr. Gun Lee who can be contacted at gun.lee@canterbury.ac.nz. He will be pleased to discuss any concerns you may have about participation in the project.

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee, and participants should address any complaints to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

If you agree to participate in the study, you are asked to complete the consent form and return back to the researcher prior to the commencement of the experiment.

Sathya Kumar Barathan
Masters Student
HITLab NZ, University of Canterbury

D

Human Interface Technology Laboratory New Zealand
Telephone: +64 3 364 2349
Email: sathya.barathan@pg.canterbury.ac.nz

Generating an efficient remote collaboration environment using shared gaze

Consent Form for participants

- ☐ I have been given a full explanation of this project and have had the opportunity to ask questions.
- ☐ I understand what is required of me if I agree to take part in the research.
- ☐ I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.
- ☐ I understand that any information or opinions I provide will be kept confidential to the researcher and that any published or reported results will not identify the participants or their institution. I understand that a thesis is a public document and will be available through the UC Library.
- ☐ I understand that all data collected for the study will be kept in locked and secure facilities and/or in password protected electronic form and will be destroyed after five years.
- ☐ I understand the risks associated with taking part and how they will be managed.
- ☐ I understand that I am able to receive a report on the findings of the study by contacting the researcher at the conclusion of the project.
- ☐ I understand that I can contact the researcher, Sathya Kumar Barathan (sathya.barathan@pg.canterbury.ac.nz) or supervisor, Gun Lee (gun.lee@canterbury.ac.nz) for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz)
- ☐ By signing below, I agree to participate in this research project.

Name:____________________________ Signed:__________________________ Date:_______________

E

Pre-experiment questionnaire (Please fill out before you start your task)

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

---

1. Your gender - Male / Female (Please circle the answer)

2. How old are you? ______________ years old

3. Relationship with your experiment partner? ______________

   ☐ Family

   ☐ Friend

   ☐ Colleague (working relationship)

   ☐ Other: ______________

4. Have you ever used a video conferencing system, such as Skype or Google Hangout?

   YES / NO (Please circle the answer)

   If yes, how often do you use the video conferencing system?

   ____

   ☐ Everyday

   ☐ Few times a week

   ☐ Few time a month

   ☐ Few times a year

   ☐ Once a few years

5. Whom do you usually have a video conferencing with?

   ____

Questionnaire for remote users after completion of Condition ___

UC UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

	Strongly disagree 1	2	3	4	Strongly agree 5
1. I think that I would like to use this system frequently					
2. I found this system unnecessarily complex					
3. I thought this system was easy to us					
4. I think that I would need the support of a technical person to be able to use this system					
5. I found the various functions in this system were well integrated					
6. I thought there was too much inconsistency in this system					
7. I would imagine that most people would learn to use this system very quickly					
8. I found the system very cumbersome to use					
9. I felt very confident using the system					
10. I needed to learn a lot of things before I could get going with this system					

Feedback on participant's collaborative experience for Condition ___

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

	Strongly disagree 1	2	3	4	Strongly agree 5
1. I noticed (my partner)					
2. (My partner) noticed me					
3. (My partner's) presence was obvious to me.					
4. My presence was obvious to (my partner).					
5. (My partner) caught my attention					
6. I caught (my partner's) attention.					
7. I was easily distracted from (my partner) when other things were going on.					
8. (My partner) was easily distracted from me when other things were going on.					
9. I remained focused on (my partner) throughout our interaction.					
10. (My partner) remained focused on me throughout the interaction					

	Strongly disagree				Strongly agree
	1	2	3	4	5
11. (My partner) did not receive my full attention.					
12. I did not receive (my partner's) full attention.					
13. My thoughts were clear to (my partner).					
14. (My partner's) thoughts were clear to me.					
15. It was easy to understand (my partner).					
16. (My partner) found it easy to understand me.					
17. Understanding (my partner) was difficult.					
18. (My partner) had difficulty understanding me.					
19. I could tell how (my partner) felt.					
20. (My partner) could tell how I felt.					
21. (My partner's) emotions were not clear to me.					
22. My emotions were not clear to (my partner).					
23. I could describe (my partner's) feelings accurately.					
24. (My partner) could describe my feelings accurately.					

Post-experiment questionnaire (fill out after you have completed all conditions)

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

1. Please rank the conditions based on your preference of usage

   *Conditions: **V**oice only, **M**ouse controlled cues, **G**aze Cues*

Rank the conditions (1,2,3) *(1 - best, 3 – worst)*	**V**	**M**	**G**
At helping you **complete** the task			
At making you feel **connected** with your partner			
At helping you stay **focused** on the task			
At making you feel that you were **present** with you partner			
For you (or the partner) to **know** that the partner (or you) **needed assistance**?			
At helping you **understand** the partner's message?			

2. What did you like the most in the condition that you ranked best?

3. What did you dislike in the condition that you ranked least?

4. Please state what could be improved in the 3 conditions.

5. Any further comments…

Thank you for your participation!

F

Verbal script to be followed during the experiment

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Welcome to the Human Interface Technology lab, New Zealand. Thank you for accepting to participate in my user research. *(Let the participants sit down in their respective positions)*

The purpose of this study is to identify and understand the efficiency of different cues during the process of remote collaboration. In this experiment, I request you to work with each other on the control panel displayed before you *(point towards the vision space screen).*

Please be reminded that we will also be recording the experiment for further analysis. However, none of your faces would be focussed. It is just to double check that the task has been carried out as per instructions.

The results will be published in academic publications including my Masters' Thesis, but all the results and collected information will be anonymised to ensure privacy.

Overall process of the experiment would be between 45 minutes – 1 hour.

Kindly read the information sheet and the consent form about the details. If you are happy with the mentioned process, please sign the consent form. Please do let me know if you have any further questions *(Let the participants read and sign the consent form)*

Before starting the experiment, please fill in the pre-experiment questionnaire for me to understand your usage of video conferencing systems and collect your demographic information.

Assign the roles to the participants (instructor/remote user and helper/local user)

To the instructor/remote user
Let's start calibrating your gaze with the eye tracker in front of you.

After calibration (running the calibration twice to ensure accuracy)
You would be guiding the local user to carry out the tasks assigned to you.

Brief the local user about the Head-mounted device

Run a couple of pilot tests for the participants to get comfortable with the devices.

I just want to investigate on how efficient different cues are during a remote collaborative process. Please remember that there is nothing right or wrong with what you do. Just go with the flow of carrying out the task together. Hopefully it would be insightful and fun.
(Start recording)

The instructor would be given a list of tasks (press buttons, turn off/on, etc.) that he/she would be guiding the local user to carry out on the large vision space screen. Let them play around with it at will.

After each condition, both the participants would be requested to fill out their feedback on how they felt about the collaborative process.

Make sure all the questions have been answered.

After they have given their feedback on all three conditions, give out the collaborative feedback questionnaire to be answered by both participants separately.

Thank you for your time and patience in answering all the questions. I request you to give me a few qualitative feedback on the entire experience of using my system.

Finally, request them to complete the simple post-experiment survey.

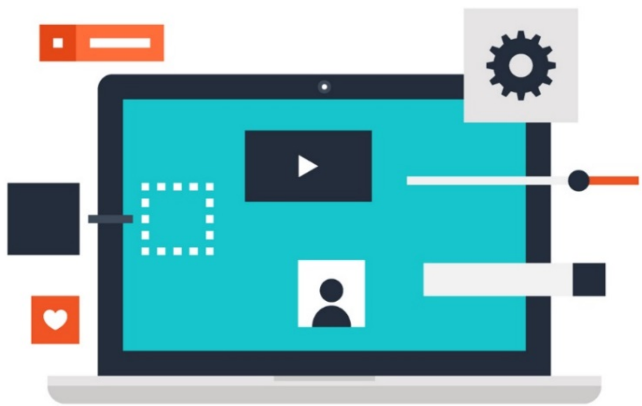
Do you have any further questions that you would like me to answer?

After the interaction, give out their vouchers.

Thank you for your participation. Enjoy your day!

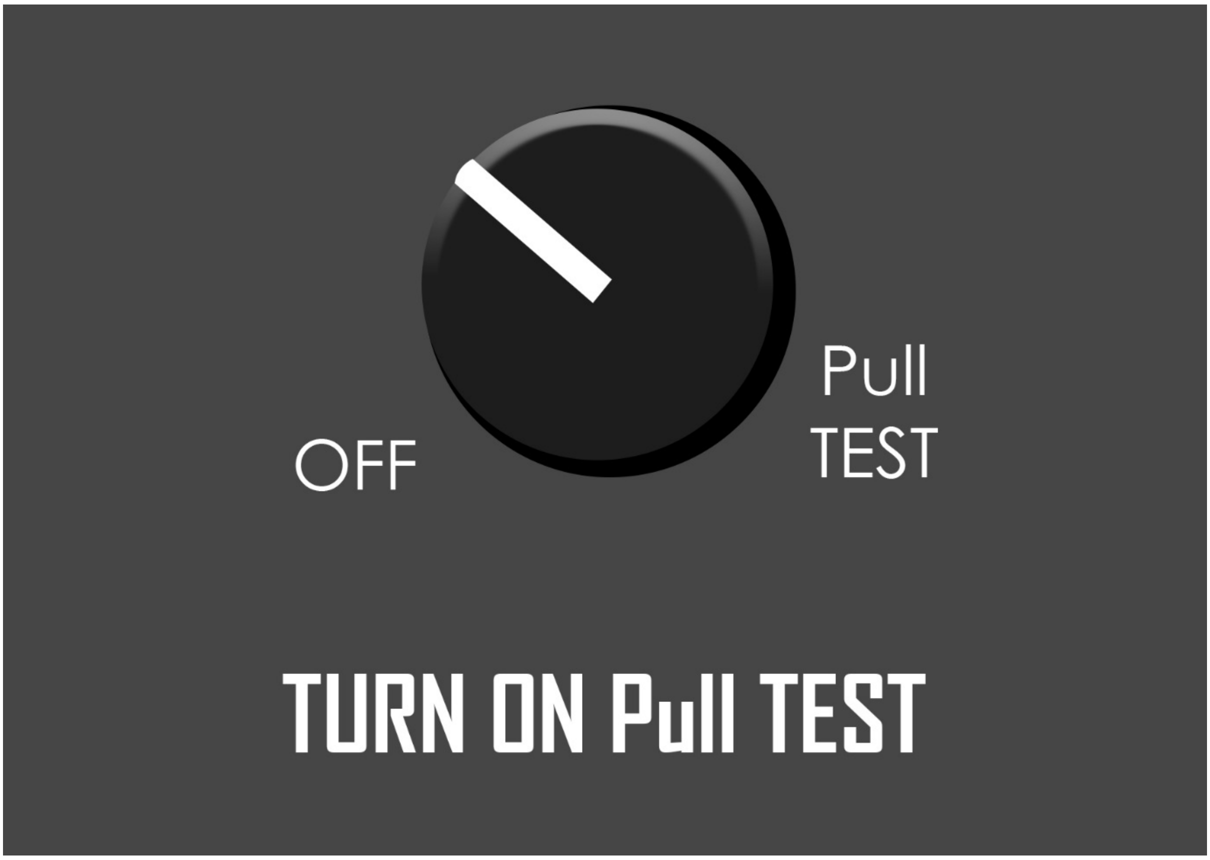
G

TURN ON SWITCH



TURN ON SWITCH


TURN ON SWITCH


OFF
Pull
TEST
TURN ON Pull TEST

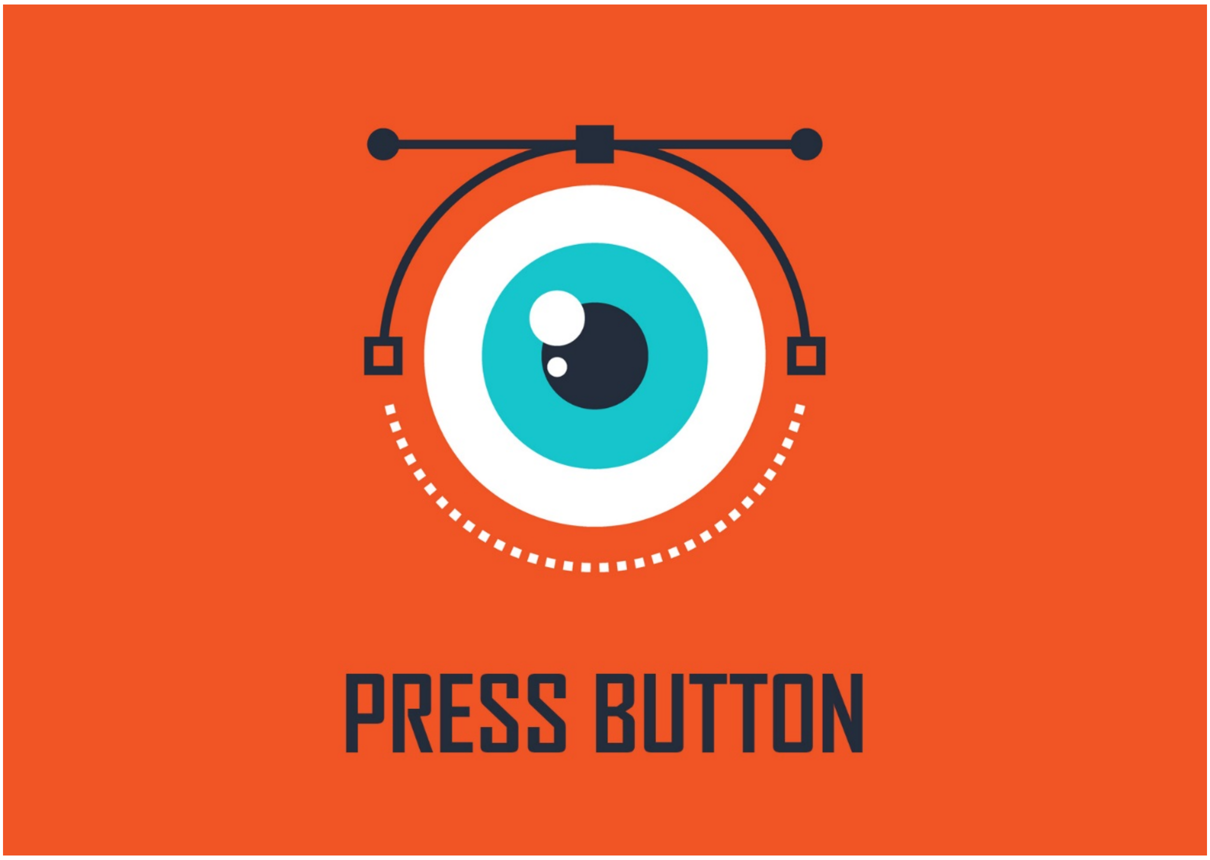
Pilot

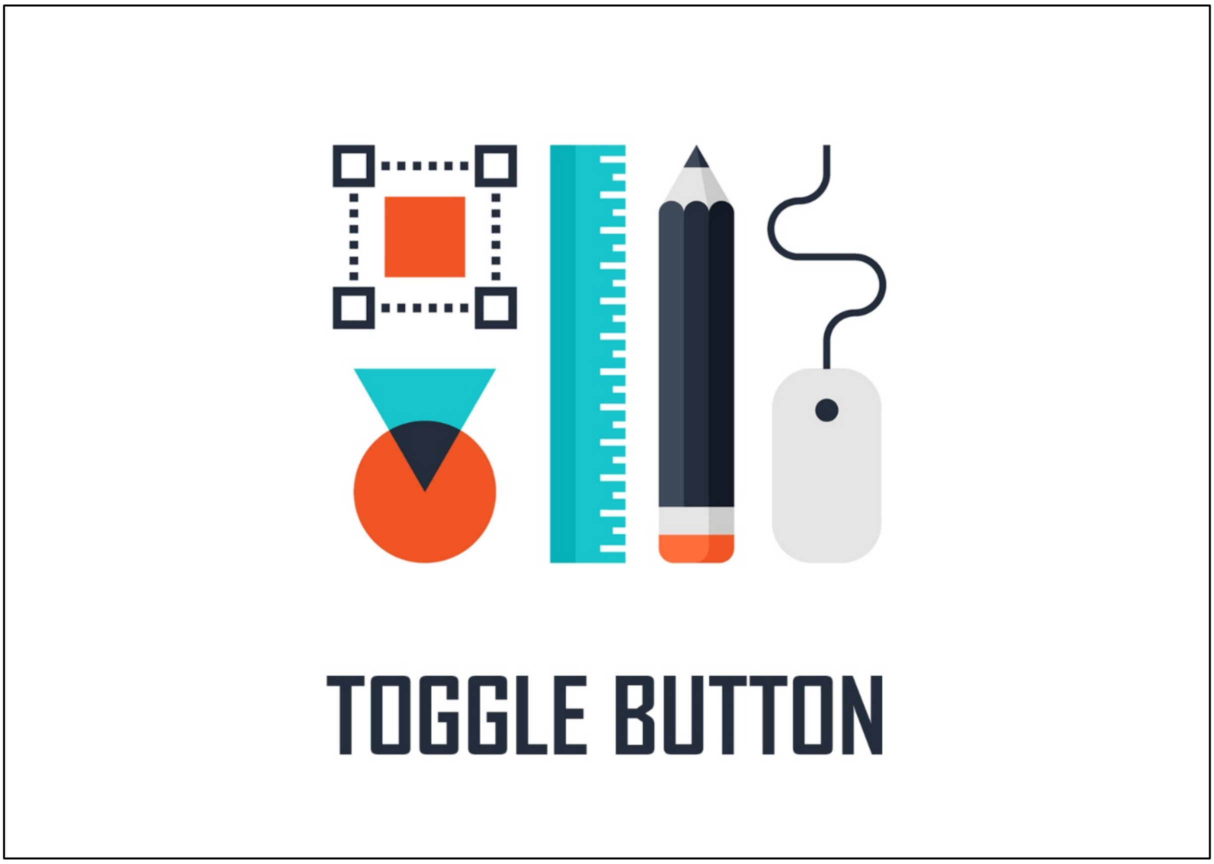
TURN SLIDER RIGHT

MOVE SLIDER TO MAX


PRESS BUTTON

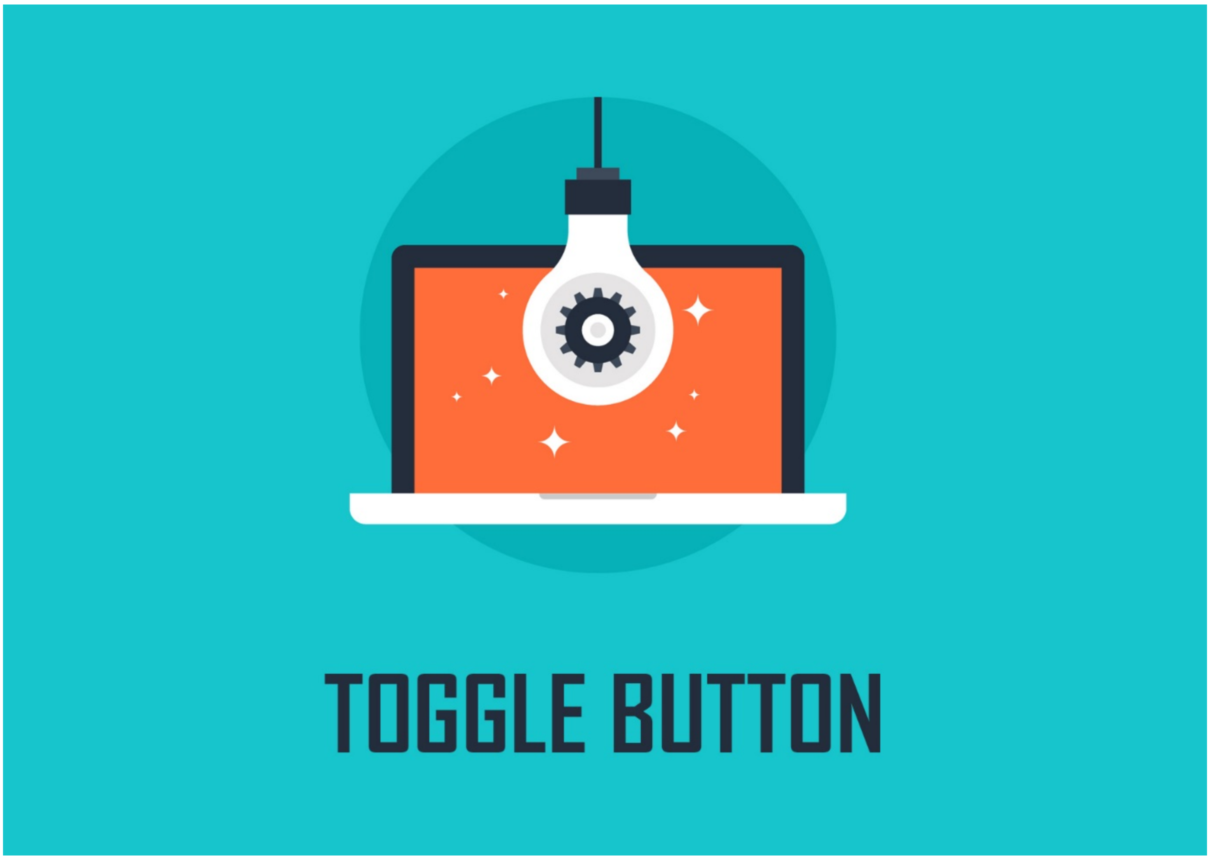

PRESS BUTTON


PRESS BUTTON


TOGGLE BUTTON

PRESS THE BUTTON

TURN OFF SWITCH


TOGGLE BUTTON


PRESS BUTTON

MOVE SLIDER TO MIN